Path integral approaches to subnetwork description and inference

Bravi, Barbara

Awarding institution:
King's College London

Thesis submitted to King's College London for the degree of Doctor of Philosophy

Path integral approaches to subnetwork description and inference

Candidate: Barbara Bravi

Supervisor: Prof. Peter Sollich

Abstract

Path integral formalism is a powerful tool borrowed from theoretical physics to build dynamical descriptions, yet its potential is largely unexplored in the context of complex networks, such as the ones common in systems biology. In this PhD thesis, I present different mathematical frameworks based on path integrals to capture the time evolution of interacting continuous degrees of freedom, e.g. biochemical concentrations. The generality of path integral approaches enables us to tackle several questions related to modelling and inference for dynamics. We first develop a novel mean field approximation, the Extended Plefka Expansion, for stochastic differential equations exhibiting generic nonlinearities. The key element is the definition of "effective" fields which map an interacting dynamics into the "most similar" non-interacting picture, i.e. the one producing the same average observables. In the resulting picture, couplings between variables are replaced by a memory and a coloured noise. We next apply this setup to the case in which part of the network is observed and part is unknown. We study the accuracy of prediction of the unobserved dynamics as a function of the number of observed nodes and other structural parameters of the system. The Extended Plefka Expansion is expected to become exact in the limit of infinite size networks with couplings of mean field type, i.e. weak and long-ranged. We show this explicitly for a linear dynamics by comparison with other methods relying on Random Matrix Theory. We finally appeal to path integrals to design "reduced" models, where equations are referred solely to some selected variables (subnetwork) but still carry information on the whole network. This model reduction strategy leads to substantially higher quantitative accuracy in the prediction of subnetwork dynamics, as we demonstrate with an example from the protein interaction network around the Epidermal Growth Factor Receptor.

Denn um dem Denken eine Grenze zu ziehen, müssten wir beide Seiten dieser Grenze denken können (wir müssten also denken können, was sich nicht denken lässt). Die Grenze wird also nur in der Sprache gezogen werden können und was jenseits der Grenze liegt, wird einfach Unsinn sein. [...]
Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.[1]

Ludwig Wittgenstein, "Tractatus Logico-Philosophicus"

This model will be a simplification and an idealization, and consequently a falsification. It is to be hoped that the features retained for discussion are those of greatest importance in the present state of knowledge.

Alan Turing, "The chemical basis of morphogenesis"

[1] In order to draw a limit to thinking we should have to be able to think both sides of this limit (we should therefore have to be able to think what cannot be thought). The limit can, therefore, only be drawn in language and what lies on the other side of the limit will be simply nonsense. [...] The limits of my language mean the limits of my world.

Acknowledgements

First and foremost I express my gratitude to Prof. Peter Sollich for his high-level, attentive and patient supervision. In this same regard I want to thank Prof. Manfred Opper, Prof. Olivier Martin and Prof. Giuseppe Longo for their guidance during my secondments at Technical University Berlin and at University Paris 11, Orsay. Their way of conceiving and carrying out novel research projects is an invaluable, formidable example to learn and to follow. The financial support from a Marie Curie Initial Training Network, NETADIS (FP7, grant 290038), is also gratefully acknowledged. I am indebted especially to all the people involved in NETADIS, as well as in the Disordered Systems group at King's College London, who enormously contributed to create a stimulating and dynamic research environment. Among them, I would like to thank in particular Reimer Kühn and Luca Dall'Asta for insightful discussions, Ludovica Bachschmid-Romano and Silvia Grigolon for their helpful, friendly and inspirational approach to our collaboration in Berlin and Paris.

Contents

1 Introduction **1**

1.1 Complexity, Model reduction and Statistical approximations 1

1.2 Overview 3

2 Extended Plefka Expansion **6**

2.1 Introduction 6

2.2 Plefka Expansion 8

2.3 Extended Plefka Expansion 11

2.3.1 Structure of the non-interacting problem 15

2.3.2 First order: Mean Field equations 17

2.3.3 Second order: TAP equations 19

2.3.4 Linear case 22

2.4 Exactness in the thermodynamic limit 23

2.4.1 Motivation and setup 23

2.4.2 Extended Plefka Expansion 25

2.4.3 Exact Solution 26

2.5 Quantitative results 31

2.5.1 Existence of poles 32

2.5.2 Terminal decay rate of correlation function 33

2.5.3 Power Spectra and Power Laws 36

2.5.4 Time Domain 40

2.6 Discussion and Conclusion 46

A Residue calculation **49**

A.1 Residue calculation 49

B Inverse Laplace transform **57**

B.1 Inverse Laplace transform 57

C Complete TAP equations **61**

C.1 Complete TAP equations 61

C.1.1 p-spin model 64

3 Inference for dynamics: the Extended Plefka Expansion with hidden nodes **69**

3.1 Introduction 69

3.2 Extended Plefka Expansion with hidden nodes 71

3.2.1 Posterior Means 78

3.2.2 Posterior Variance 80

3.2.3 Thermodynamic Limit 82

3.2.4 Comparison with known results 84

3.3 Power spectrum and critical regions 85

3.3.1 Dimensionless system 85

3.3.2 Critical scaling 86

3.3.3 Master curves for $\gamma \to \gamma_c$ and $\alpha \to 0$ 88

3.3.4 Master curves for $\alpha \to 1$ and $p \to 0$ 92

3.4 Posterior variance in the time domain 94

3.4.1 Relaxation time for $\gamma \to \gamma_c$ and $\alpha \to 0$ 94

3.4.2 Relaxation time for $\alpha \to 1$ and $p \to 0$ 99

3.4.3 Equal time posterior variance for $\gamma \to \gamma_c$ and $\alpha \to 0$ 102

3.4.4 Equal time posterior variance for $\alpha \to 1$ and $p \to 0$ 103

3.5 Discussion and Conclusion 106

D Discrete time operator inversion **109**

D.1 Discrete time operator inversion 109

4 Inference for dynamics: exact average case performance 112

4.1 Introduction 112

4.2 Model and general expression for posterior covariance 115

4.3 Thermodynamic limit by Random Matrix Theory 119

4.3.1 Feedback matrix: Wishart ensemble 119

4.3.2 Self-interacting hidden variables 120

4.3.3 Symmetric hidden-hidden couplings 125

4.4 Thermodynamic Limit by Dynamical Functionals 132

4.4.1 Asymmetric hidden-hidden couplings 134

4.4.2 Generalization to arbitrary interaction symmetry 137

4.5 Discussion and Conclusion 139

E Kalman filter and smoother 143

E.1 Kalman filter and smoother 143

F Variational method 149

F.1 Variational method 149

5 Gaussian Variational Approximation for biochemical networks 152

5.1 Introduction 152

5.2 Set up 155

5.2.1 Projection Methods 157

5.3 Gaussian Variational Approximation 159

5.4 Subnetwork effective description within GVA 162

5.4.1 Bulk and subnetwork 162

5.5 Nonlinear corrections by perturbation theory 166

5.5.1 Nonlinear memory 166

5.5.2 Numerical implementation 170

5.5.3 Comparison with projection methods . . . 171
5.6 Application to biochemical networks . . . 171
5.6.1 Toy model . . . 172
5.6.2 Application to EGFR . . . 175
5.7 Discussion and conclusion . . . 181

G Gaussian Variational Approximation **188**
G.1 Gaussian Variational Approximation . . . 188

H Memory and coloured noise in the linearized dynamics **195**
H.1 Stationary case . . . 195
H.2 Generalization to non-stationary case . . . 198

I Nonlinear corrections by perturbation theory **199**
I.1 Conditional Gaussian moments . . . 199
I.1.1 Comparison with the Extended Plefka Expansion and the Kalman Filter . 201
I.2 Perturbative expansion . . . 203

J Projection methods vs GVA **207**
J.1 Projection methods vs GVA . . . 207
J.2 Linearized projected equations . . . 208
J.3 Nonlinear projected equations . . . 209
J.4 Full nonlinear memory and random force in the GVA . . . 212
J.5 Proof of the equivalence . . . 215
J.5.1 Linear order in δ . . . 215
J.5.2 Quadratic order in δ . . . 215

K Effective equations solver **219**
K.1 Effective equations solver . . . 219

6 Summary and Future directions **221**

6.1 Summary 221

6.2 Future directions 224

6.2.1 Implementation and performance 224

6.2.2 Coarse-graining fluctuations in gene expression 225

6.2.3 Learning rules and experiment design 226

6.2.4 Connection to information-theoretic tools 228

References **230**

List of Figures

2.1 $\lambda_{\text{threshold}}$ and $\lambda_{\min}$ as a function of η. 33

2.2 Typical plot for $\lambda > 2$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. . 34

2.3 Typical plot for $8/3\sqrt{3} < \lambda < 2$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. 35

2.4 Typical plot for $\lambda < 8/3\sqrt{3}$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. 35

2.5 Normalized power spectra for different symmetries 37

2.6 Power spectrum at $\lambda = 1 + \eta$ for positive η 38

2.7 Power spectrum at $\lambda = 1 + \eta$ for negative η 38

2.8 Correlations in temporal domain for positive and negative symmetry 41

2.9 Temporal correlations and asymptotic curves for $\eta = 1$. 43

2.10 Temporal correlations and asymptotic curves for $\eta = -1$. 43

A.1 Singularities of $\tilde{C}(z)$ in the complex z-plane for $\eta = 0$ 50

A.2 Singularities of $\tilde{C}(z)$ for $\eta > 0$. 51

A.3 Region of the complex plane where $\tilde{C}(z)$ for $\eta = 1$ is defined. 51

A.4 Singularities of $\tilde{C}(z)$ for $\eta < 0$. 52

A.5 Region of the complex plane where $\tilde{C}(z)$ for $\eta = -1$ is defined. 52

A.6 Integration contour for $\eta = 0$. 53

A.7 Integration contour for I_1 for $\eta > 0$. 55

A.8 Integration contour for I_2 for $\eta > 0$. 55

A.9 Integration contour for I_1 for $\eta < 0$. 56

A.10 Integration contour for I_2 for $\eta < 0$. 56

B.1 Integration contour for the inverse double sided Laplace transform of $\tilde{C}(z)$. 58

B.2 Subtraction of $g_{\text{pole}}(\omega)$ from $\tilde{C}(\omega)$ for $\epsilon = 0.05$ 60

B.3 Numerical inverse Laplace transform at $\epsilon = 0.05$. 60

3.1 Parameter space spanned by α, γ, p. 88

3.2 Numerical solutions for small α and for small $\delta\alpha = \alpha - 1$. 95

3.3 Master curves for different values of $\bar{\epsilon}$ and η. 96

3.4 $C(\Omega)$ for different α, at small fixed η. 97

3.5 Rescaled relaxation time $\bar{\tau}(\bar{\alpha})$ and relaxation time τ for $\alpha \to 0$ and $\gamma \to \gamma_c$. . . . 100

3.6 Relaxation time for $\alpha \to 1$ and $p \to 0$. 101

3.7 Inference error for $\alpha \to 0$, $\gamma \to \gamma_c$, $p \to 0$ as a function of α. 104

3.8 Inference error for $\alpha \to 1$ and $p \to 0$, as a function of α. 105

4.1 $\rho(\tau)$ for $\alpha = 0.5$ and $\rho(\ln\tau)$ for different α at $j = 0$. 123

4.2 Posterior relaxation times and posterior power spectrum for various α at $j = 0$. . . 124

4.3 $\rho(\ln\tau)$ at $\eta = 1$ for various α. 128

4.4 Two pieces spectrum $\rho(\ln\tau)$ for $\eta = 1$. 130

E.1 Linear-Gaussian state space model. 144

5.1 Time courses and error of approximation for the toy model. 176

5.2 EGFR reaction network. 176

5.3 Time courses and error of approximation for the EGFR network. 177

5.4 Nonlinear self-memory of subnetwork species SOS. 179

5.5 Nonlinear cross-memory of SOS for the product GS-SOS. 180

5.6 GVA-edge and GVA-interior terms in the entire nonlinear memory of SOS. . . . 180

Chapter 1

Introduction

> La explicación es obvia: *El jardín de los senderos que se bifurcan* es una imagen incompleta, pero no falsa, del universo tal como lo concebía Ts'ui Pên. [...] Creía en infinitas series de tiempos, en una red creciente y vertiginosa de tiempos divergentes, convergentes y paralelos. Esa trama de tiempos que se aproximan, se bifurcan, se cortan o que secularmente se ignoran, abarca *todas* la posibilidades.[1]
>
> *Jorge Luis Borges, "El jardín de senderos que se bifurcan"*

1.1 Complexity, Model reduction and Statistical approximations

Many physical systems involve the interaction of numerous components and exhibit a range of what we can call *complex* behaviours, for instance a *nonlinear*, *stochastic*, *out-of-equilibrium* dynamics. Systems biology typically aims at describing complex systems of this kind, and a remarkable example are the protein-protein interaction networks building up signalling pathways: they can be of extremely large size - with thousands of reacting species - and the mechanisms underlying their functioning are still substantially unclear. Both fundamental and practical limitations have emerged in quantitative approaches to studying these systems. Nonlinearities are hardly tractable, we still lack tools to exhaustively characterize non-equilibrium and stochastic regimes and the combinatorial increase of the number of possible connections with interacting partners leads to a vast amount of information being encoded in biological networks. The comparison itself with experimental results can be subject to an overall *uncertainty* due to missing variables and

[1] The explanation is obvious: *The Garden of Forking Paths* is an incomplete, but not false, image of the universe as Ts'ui Pên conceived it. [...] He believed in an infinite series of times, in a growing, dizzying net of divergent, convergent and parallel times. This network of times which approached one another, forked, broke off, or were unaware of one another for centuries, embraces *all* possibilities of time.

low-resolution measurements. Even if there was the possibility of accessing the entire network, a complete qualitative understanding of the kinetics along multiple pathways might not be achieved. Modelling and data analysis in systems biology are therefore confronted with major challenges. First, the dimensionality of the phase space and the presence of nonlinearities in the temporal evolution are such that we need to resort to *approximate* descriptions and *model reduction* techniques [1, 2]. The latter are designed to reduce a problem to equations for fewer variables with the minimal loss of mesoscopic detail, i.e. in such a way that the solution still retains dynamical features of the full set of equations. Some methods for example are based on the exclusion of fast variables [3, 4] or of the ones with a negligible impact on the output behaviour [5]; other model reduction strategies proceed by "lumping" together, rather than removing, the elements with similar dynamical features [6, 7]. The aim is to simplify the overall analysis, to facilitate the interpretation of experimental data and ultimately to extract useful information from them. Model reduction can be put forward as a tool for *simplification* and *better understanding* in complex systems such as biochemical pathways.

As a second challenge, the interface between models and data in stochastic, non-equilibrium settings is an important problem to address, to quantify the range of uncertainty by which inferences we draw on the system are affected. The term *inference* here is intended as the use of available information for updating our knowledge on unknown states and/or parameters.

The need for including stochastic effects in biology has been acknowledged in several contexts [8, 9]; the Chemical Master Equation (CME) [10] describes rigorously stochastic biochemical reactions in terms of probability distributions but no analytic solution is known, therefore their investigation relies on simulation approaches such as the Stochastic Simulation Algorithm (SSA), a Monte Carlo method which samples from the solutions of CME [11, 12]. These are fundamental tools, yet they often require in parallel, to be implemented in a less computationally expensive way, some clever simplification based on model reduction prescriptions. In this regard, improvements over the SSA are given by mixed methods: they exploit either timescales separation, i.e. the fact that some reactions occur on fast scales and others on slow ones (see e.g. the finite-state projection [13, 14]), or similar separation in the abundance of different reactants [15]. In addition, in simulation approaches, it is not clear how to subsume available observations in a *systematic* and *principled* way as constraints on the dynamics under investigation. *Analytic* approximations of probability distributions can be seen as an extremely valuable tool in this respect, while giving further insight into the underlying processes. Two particularly powerful approaches to derive

approximate descriptions for stochastic as well as "disordered"[2] systems, both in the statistical physics and machine learning community, are variational techniques and weak coupling expansions: these will be the core ideas of the work we present here.

Furthermore, equilibrium systems are well studied, both from the modelling and the inference point of view, as efficient approximations and algorithmic techniques have been developed [16]. On the other hand, much less has been done for systems which exhibit non-equilibrium steady states, whereas they deserve particular attention in systems biology. Very often in fact the dynamics is subject to fluxes driving it out of equilibrium, e.g. in typical reaction-diffusion processes. When dealing with these, it is convenient to choose as a starting point tools incorporating the *dynamical* information on the whole temporal trajectory. Exactly the need to deal with a variety of non-equilibrium scenarios leads us to introduce probability distributions in the *path* (i.e. temporal trajectories) space, which contain information on both transient and steady state behaviours as well as all the time-dependent observables. In a few words, path integrals allow one to interpret probabilistic quantities as a sum over all possible ways of evolution, or histories, of the system. This idea can be traced back to the theory of diffusion by Wiener [17], it has since been applied via different formulations to stochastic processes [18–23] and recently has received attention in biological modelling, from neuroscience [24] to polymer physics [25].

1.2 Overview

The aim of this PhD thesis is to investigate how path integral approaches can be tailored to the tasks of mean field analysis, model reduction and optimal state prediction. Path integral representations of stochastic dynamics are the unifying thread among chapters, with Langevin equations for continuous degrees of freedom being the common starting point. Given the wide range of tools used in the different parts of this thesis, we provide more specific details in the introduction of each chapter. Similarly we add an overview and a discussion for each set of results, which is conceived as a stand-alone piece of original work, published or to be published autonomously.

The thesis is organized as follows. In chapter 2, we propose an extension of the Plefka expansion, which is well known for the dynamics of discrete spins, to stochastic dynamics with continuous degrees of freedom and exhibiting generic nonlinearities. The main feature of our approach is to constrain in the Plefka expansion not just first moments akin to magnetizations, but also second moments, specifically two-time correlations and responses for each degree of freedom, resulting

[2] In the standard usage "disorder" refers to a quenched ("frozen") stochasticity.

in a Gaussian representation of the dynamical trajectories. The end result is an effective equation of motion for each single degree of freedom, where couplings to other variables appear as a self-coupling to the past (i.e. memory term) and a coloured noise. Such a single-variable representation relies on the definition of "effective" fields which yield by construction an average characterization equivalent to the original interacting dynamics while making the implementation more convenient computationally. The extended Plefka expansion constitutes a new mean field approximation that should become exact in the thermodynamic limit of a large network, for suitably long-ranged couplings. For the analytically tractable case of linear dynamics we establish this exactness explicitly by appeal to spectral methods of Random Matrix Theory, for Gaussian couplings with arbitrary degree of symmetry (thus capturing also out-of-equilibrium systems).

In chapter 3, we generalize the Extended Plefka Expansion to the case of just partially observed networks, which could be relevant for inference methods and experiment design. We consider the problem of a subnetwork of observed nodes embedded into a larger bulk of unknown (i.e. hidden) nodes, where the aim is to infer these hidden states given information about the subnetwork dynamics. As a paradigmatic model we study the stochastic linear dynamics of continuous degrees of freedom interacting via random Gaussian couplings. The resulting posterior distribution over the hidden dynamics is known to be Gaussian and this allows us to fully characterize it in terms of the first two moments conditional on the observed trajectories. While the posterior mean gives the best estimate of the hidden dynamics, the equal-time posterior variance gives the expected error of this prediction. We study this error in the infinite network size limit and the stationary regime by resorting to the Extended Plefka Expansion. We analyze its phase diagram in the space of the system parameters and the ratio α between the number of hidden and observed nodes. In particular, we identify critical regions in this parameter space where the state prediction error diverges, and determine the corresponding scaling behaviour in their proximity.

In chapter 4, we consider again average performance results for dynamical inference problems in large networks, where a set of nodes is hidden while the time trajectories of the others are observed. Similarly to chapter 3, we focus on the linear stochastic dynamics of continuous variables interacting via random Gaussian couplings of generic symmetry and analyze the prediction error. By applying Kalman filter recursions we find that the posterior dynamics is governed by an "effective" drift that incorporates the effect of the observations. We present two approaches for characterizing the posterior variance that allow us to tackle, respectively, equilibrium and non-equilibrium dynamics. The first appeals to Random Matrix Theory and reveals average spectral properties of the inference error and typical posterior relaxation times, the second is based on dynamical func-

tionals and yields the inference error as the solution of an algebraic equation. We successfully compare these exact results with the ones in chapter 3, derived by appeal to the Extended Plefka Expansion with hidden nodes in the thermodynamic limit of infinitely large system.

As an effectively non-interacting picture, the Extended Plefka Expansion has to neglect some fundamental dynamical information on the coupling between parts of the network, which might become negligible for large networks but is crucial for a realistic description of *small* systems. Therefore in chapter 5, we introduce another path integral approach aimed at a Gaussian representation of stochastic dynamics without resorting to a mean field approximation: the aim is to produce descriptions more adequate for the microscopic scale and for the detailed dynamical analysis of subnetworks.

In more detail, we apply a Gaussian variational approximation to model reduction in networks of unary and binary biochemical reactions that are partially accessible experimentally. In the spirit of model reduction, we restrict the analysis to a small subset of variables (subnetwork), embedded in a larger network (bulk) and the key goal is to write dynamical equations reduced to the subnetwork but still retaining the feedback from the bulk. As a result, the subnetwork reduced dynamics contains a memory and an additional stochastic term exhibiting non trivial temporal correlations (a coloured noise). We first derive the subnetwork equation for the linearized (Gaussian) dynamics; next a perturbative power expansion allows us to estimate first order nonlinear corrections. If one specializes to the case of vanishing intrinsic noise, our description is explicitly shown to be equivalent to projection methods up to quadratic terms, yet it is applicable more generally in the presence of stochastic fluctuations. An example from the Epidermal Growth Factor Receptor pathway, relevant to cancer signalling, is also provided to probe the increased accuracy of prediction and the computational convenience of our method.

Finally, in chapter 6, we summarize all these results, we draw conclusions on their significance and potential applications and we sketch some directions for future work.

CHAPTER 2

Extended Plefka Expansion

This work has been done in collaboration with Manfred Opper (TUB Berlin). The related paper is published in Journal of Physics A: Mathematical and Theoretical [26].

> Automation brings in real "mass production" not in terms of size, but of an instant inclusive embrace. Such is also the character of "mass media". They are an indication, not of the size of their audiences, but of the fact that everybody becomes involved in them at the same time.
>
> *Marshall McLuhan, "Understanding Media"*

2.1 Introduction

Stochastic Differential Equations (SDEs) with continuous variables are a well-established tool to describe the dynamical behaviour of a variety of systems, in areas ranging from physics and chemistry to biology and engineering [10, 27]: they are used frequently, for example, for dynamical modelling of intracellular kinetics and biochemical networks [28].

In the context of network studies, in particular with regard to applications in systems biology, a major task is model simplification [5, 6], using model reduction strategies that should retain as much as possible of the qualitative dynamical information - as we briefly discussed in chapter 1. In addition one requires techniques amenable to the inference of model parameters from observed data, since experimental uncertainties on parameters, resulting e.g. from the fact that some dynamical variables may not be observed, can crucially affect the predictions of dynamical models [29].

The application of approaches based on statistical mechanics and spin glass theory has a long

history [30]. In particular, Mean Field (MF) methods have emerged as powerful tools for characterizing statistical quantities in systems where the combinatorial complexity of exact calculations rules out a tractable description [31]. From the theoretical point of view, further motivation for the use of mean field methods comes from the fact that they can often be proved to retrieve the exact solution in an appropriate limit, typically involving high network connectivity and/or weak couplings.

The so called “Plefka expansion" for the Sherrington-Kirkpatrick (SK) [32] model was introduced by Plefka [33] as a convenient method to derive MF equations and their more refined analogue, the TAP equations [34]. The advantage of the method, essentially an expansion of the Gibbs free energy in powers of the interaction strength, is that it does not rely on an average over interactions drawn from some statistical ensemble. This makes it potentially useful in applications to e.g. biology, where it is generally a *specific* network that is of interest.

Roudi and Hertz [35] applied the Plefka expansion to the problem of approximating spin-glass dynamics: in this case, variables are not single spins but entire time histories of each spin. They developed a dynamical theory that relates mean magnetizations, potentially time varying fields and quenched couplings for two versions of SK model kinetics (synchronous and asynchronous updates, respectively). Using the generating functional approach, the (naive) MF and TAP dynamical equations were retrieved as first and second orders of a power expansion in analogy with the equilibrium Plefka expansion for the Gibbs free energy. In more detail, the logarithm of the generating functional for the dynamics plays the role of the equilibrium free energy: performing the Legendre transform w.r.t. the real and auxiliary fields one obtains the dynamical equivalent of the Gibbs free energy and then can expand for weak couplings. Importantly, as long as the generating functional is by definition dependent only on fields that act linearly on the degrees of freedom, this expansion will closely resemble the standard Plefka approach and only the first moments of the resulting probability measure over trajectories will be fixed.

The aim of this chapter is two-fold. First we want to introduce an improvement, tailored to continuous degrees of freedom, of the approximation strategy outlined above; we call the improved method an “extended" Plefka expansion. The dynamical model is a set of stochastic differential equations for continuous degrees of freedom and with generic nonlinear couplings between them. The basic idea of the extension that we propose is to include among the set of order parameters all second moments, i.e. two-time correlations and responses, for each degree of freedom. Expanding up to second order in interaction strength then provides a mean field description where couplings between trajectories are replaced by a coupling to the past (i.e. a memory term) and a coloured

noise.

Our second aim is an analytical investigation of a solvable limit, which concerns large networks with linear dynamics. This partly serves the purpose of verifying explicitly a case where the approximation becomes exact, but the calculation also provides additional insight into how the dynamical behaviour of correlations and responses depends on the symmetry of the couplings. We show that the exact thermodynamic limit is recovered from the approximate equations for any degree of symmetry, i.e. irrespective of whether the system reaches an equilibrium stationary state. This keeps the analysis as general as possible and suggests multiple possible applications, for example in neural networks and gene expression where couplings are typically asymmetric.

The chapter is organized as follows: after recalling the expansion conceived by Plefka in section 2.2, we introduce in section 2.3 the basic functional integral approach that provides the framework within which we build the extended Plefka expansion for dynamics. In sections 2.3.2 and 2.3.3 we present and discuss the derivation of the approximate dynamical equations from the functional integral. In sections 2.3.4, 2.4.1 and 2.4.2 we apply the approximation to the particular case of a linear dynamics, which is analytically tractable both in the static and dynamic scenario. In section 2.4.3 we resort to Random Matrix Theory and related spectral methods [36] to average the exact dynamics over the disordered interactions, in the limit of an infinitely large sample and in the stationary regime. This allows us to derive expressions for correlations and responses in Laplace space, and comparison with the predictions of the extended Plefka approximation shows perfect agreement. This confirms and strengthens the theoretical justification of our method. In section 2.5.3 we study in more detail the qualitative features of the dynamics, in particular non-exponential relaxation behaviour that manifests as low-frequency power law tails in the power spectra. Finally, an explicit analytical characterization of correlations and responses in the temporal domain can be found in the limit of symmetric and antisymmetric couplings and is discussed briefly in section 2.5.4.

2.2 Plefka Expansion

We briefly summarize the main steps of the "Plefka expansion" introduced by Plefka [33], using, as in the original paper, the Sherrington-Kirkpatrick (SK) [32] model as an example. The SK model of a spin glass consists of N Ising spins ($S_i = \pm 1$) with Hamiltonian

$$\mathcal{H} = \frac{1}{2}\sum_{i\neq j} J_{ij} S_i S_j + \sum_i h_i^{\text{ext}} S_i \tag{2.2.1}$$

In the SK model, specifically, the interactions are symmetric (i.e. $J_{ij} = J_{ji}$) and infinitely long-ranged, with the J_{ij} for $i < j$ chosen as independent Gaussian variables of mean zero and variance $1/N$, though these properties are not required to write down the general expansion. Note that the left hand side of (2.2.1) would conventionally be written as $-\beta\mathcal{H}$ with β the inverse temperature, but we omit factors of $-\beta$ here and below as we do not need them in the application to dynamics. In order to construct the Plefka expansion one introduces a parameter α controlling the interaction strength, defining a modified Hamiltonian as

$$\mathcal{H}_\alpha = \frac{\alpha}{2}\sum_{i\neq j} J_{ij}S_iS_j + \sum_i h_i^{\text{ext}}S_i \tag{2.2.2}$$

The full interacting Hamiltonian is then $\mathcal{H}_1 = \mathcal{H}$, while $\mathcal{H}_0$ is the Hamiltonian of a non-interacting system. The Gibbs free energy G_α is now defined as the free energy subject to a constraint on certain averages, typically the magnetizations $m_i = \langle S_i\rangle$

$$G_\alpha(\boldsymbol{m}) = \text{extr}_{\boldsymbol{h}}\hat{G}_\alpha(\boldsymbol{m},\boldsymbol{h}) \tag{2.2.3}$$

with

$$\hat{G}_\alpha(\boldsymbol{m},\boldsymbol{h}) = \ln \text{Tr}\, e^{\Xi_\alpha} \tag{2.2.4}$$

and

$$\Xi_\alpha = \mathcal{H}_\alpha + \sum_i h_i(S_i - m_i) \tag{2.2.5}$$

One can write

$$G_\alpha(\boldsymbol{m}) = \text{extr}_{\boldsymbol{h}}\left(\ln \text{Tr}\, e^{\mathcal{H}_\alpha + \sum_i h_iS_i} - \sum_i h_im_i\right) \tag{2.2.6}$$

and this shows that G_α is the Legendre transform of a Helmholtz free energy – the first term in the brackets – that depends on the auxiliary fields h_i. The extremization condition over the h_i gives

$$m_i = \langle S_i\rangle \tag{2.2.7}$$

and this ensures that the m_i have the intended meaning. The average here is over the distribution of states $P(\boldsymbol{S}) \propto e^{\Xi_\alpha}$. This is biased away from the Boltzmann distribution $(1/Z)\, e^{\mathcal{H}_\alpha}$ by the factor $e^{\boldsymbol{h}\cdot\boldsymbol{S}}$ involving the auxiliary fields h_i. We will denote the fields that produce the desired values of the magnetizations $\boldsymbol{m}$ by $\boldsymbol{h}_\alpha(\boldsymbol{m})$, where the subscript emphasizes the dependence on the interaction strength α. The fields $\boldsymbol{h}_\alpha$ can be deduced as derivatives of G_α, once this is known. Explicitly, because of the condition (2.2.7), the variation of the fields $\boldsymbol{h}_\alpha$ with $\boldsymbol{m}$ does not contribute to the $\boldsymbol{m}$-derivative of G_α, so that

$$\frac{\partial G_\alpha}{\partial m_i} = -h_{i\alpha} \tag{2.2.8}$$

as expected on general grounds from the Legendre transform definition of G_α. The Gibbs free energy becomes equal to the unconstrained equilibrium free energy when the fields $h_{i\alpha}$ vanish, so that the condition for the equilibrium magnetizations is simply

$$\frac{\partial G_\alpha}{\partial m_i} = 0 \tag{2.2.9}$$

The formalism so far is generic. In the Plefka expansion, the interacting part of the Hamiltonian is treated perturbatively by expanding the Gibbs free energy in powers of α, typically to first or second order

$$G_\alpha = G^0 + \alpha G^1 + \frac{\alpha^2}{2} G^2 + \dots \tag{2.2.10}$$

where $G^k = (\partial/\partial\alpha)^k \, G_\alpha|_{\alpha=0}$. The fields $h_{i\alpha} = -\partial G_\alpha / \partial m_i$ can be expanded analogously

$$\boldsymbol{h}_\alpha = \boldsymbol{h}^0 + \alpha \boldsymbol{h}^1 + \frac{\alpha^2}{2} \boldsymbol{h}^2 + \dots \tag{2.2.11}$$

To the second order, the equilibrium condition $\boldsymbol{h}_\alpha = 0$ for the order parameters $\boldsymbol{m}$ is then given by

$$0 = \boldsymbol{h}^0 + \alpha \boldsymbol{h}^1 + \frac{\alpha^2}{2} \boldsymbol{h}^2 \tag{2.2.12}$$

In applications to equilibrium spin systems, the non-interacting Gibbs free energy G^0 can often be found explicitly, e.g. for our Ising spin example

$$G^0 = -\sum_i \left[\frac{1+m_i}{2} \ln\left(\frac{1+m_i}{2}\right) + \frac{1-m_i}{2} \ln\left(\frac{1-m_i}{2}\right) \right] + \sum_i h_i^{\text{ext}} m_i \tag{2.2.13}$$

In dynamical problems, finding G^0 explicitly is often awkward but can be avoided by noting that in order to obtain a certain value of $\boldsymbol{m}$ at $\alpha = 0$ requires a field $\boldsymbol{h}^{\text{eff}} = \boldsymbol{h}^0$. The equilibrium condition (2.2.12) for nonzero α can then be rewritten as

$$\boldsymbol{h}^{\text{eff}} = -\alpha \boldsymbol{h}^1 - \frac{\alpha^2}{2} \boldsymbol{h}^2 \tag{2.2.14}$$

This expression gives us the effective fields $\boldsymbol{h}^{\text{eff}}$ that produce the *same* magnetizations $\boldsymbol{m}$ in the non-interacting system as at equilibrium in the interacting system. To obtain the equilibrium condition for the interacting system, one then only needs to combine this with the relation between magnetization and field in the non-interacting system, which for Ising spins reads simply

$$m_i = \tanh(h_i^{\text{ext}} + h_i^{\text{eff}}) \tag{2.2.15}$$

To carry out the actual calculation of the first and second order Plefka free energies G^1 and G^2, one notes first that $G_\alpha(\boldsymbol{m}) = \hat{G}_\alpha(\boldsymbol{m}, \boldsymbol{h}_\alpha(\boldsymbol{m}))$, hence

$$\frac{\partial G_\alpha}{\partial \alpha} = \frac{d\hat{G}_\alpha}{d\alpha} = \left\langle \frac{d\Xi_\alpha}{d\alpha} \right\rangle_\alpha \tag{2.2.16}$$

where we use $(d/d\alpha)$ to indicate a total derivative that includes the α-dependence of $\boldsymbol{h}_\alpha$. On the other hand (2.2.5) shows that in Ξ_α each field $h_{i\alpha}$ multiplies $S_i - m_i$, whose average vanishes, so this α dependence drops out and one has simply

$$\frac{\partial G_\alpha}{\partial \alpha} = \langle \mathcal{H}_{\rm int} \rangle_\alpha \tag{2.2.17}$$

where $\mathcal{H}_{\rm int} = \partial \mathcal{H}_\alpha / \partial \alpha$ is the interacting part of the original Hamiltonian. Evaluating the average in the non-interacting system ($\alpha = 0$) then gives $G^1 = \langle \mathcal{H}_{\rm int} \rangle_0$, and by derivation $\boldsymbol{h}^1$. For the SK model, one finds in this way $G^1 = (1/2) \sum_{i \neq j} J_{ij} m_i m_j$ and $h_i^1 = -\partial G^1 / \partial m_i = -\sum_{j \neq i} J_{ij} m_j$. To first order the effective field is then $h_i^{\rm eff} = -\alpha h_i^1 = \alpha \sum_{j \neq i} J_{ij} m_j$ and the equilibrium condition $m_i = \tanh(h_i^{\rm ext} + \alpha \sum_{j \neq i} J_{ij} m_j)$ has the familiar mean field form. For the second order one has in general

$$\begin{aligned} \frac{\partial^2 G_\alpha}{\partial \alpha^2} &= \frac{d^2 \hat{G}_\alpha}{d\alpha^2} = \\ &= \left\langle \frac{d^2 \Xi_\alpha}{d\alpha^2} \right\rangle_\alpha + \left\langle \left(\frac{d\Xi_\alpha}{d\alpha} \right)^2 \right\rangle_\alpha - \left\langle \frac{d\Xi_\alpha}{d\alpha} \right\rangle_\alpha^2 \end{aligned} \tag{2.2.18}$$

The first term vanishes because $\partial^2 \mathcal{H}_\alpha / \partial \alpha^2 = 0$ and because $\partial^2 h_{i\alpha} / \partial \alpha^2$ is multiplied again by a vanishing average. Evaluating at $\alpha = 0$ then gives (as discussed in [35])

$$G^2 = \left\langle \left(\delta \frac{d\Xi_\alpha}{d\alpha} \right)^2 \right\rangle_0 \tag{2.2.19}$$

where

$$\delta \frac{d\Xi_\alpha}{d\alpha} = \frac{d\Xi_\alpha}{d\alpha} - \left\langle \frac{d\Xi_\alpha}{d\alpha} \right\rangle_0 = \mathcal{H}_{\rm int} - \langle \mathcal{H}_{\rm int} \rangle_0 + \boldsymbol{h}^1 \cdot (\boldsymbol{S} - \boldsymbol{m}) \tag{2.2.20}$$

From G^2 one finds $\boldsymbol{h}^2$ by taking $\boldsymbol{m}$-derivatives again, and in principle this process can be iterated to higher order. The first order gives a MF approximation as shown above, while at second order one retrieves what are known as the TAP equations for the SK-model [32].

2.3 Extended Plefka Expansion

We start from the dynamical equations

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \phi_i(\boldsymbol{x}(t)) + \xi_i(t) \tag{2.3.1}$$

for a set of N continuous (real-valued) degrees of freedom x_i ($i = 1, \ldots, N$) evolving in time t. The x_i may represent e.g. concentrations of chemical species in a biochemical reaction network, or deviations of such concentrations from steady state values. On the r.h.s., $\phi_i(\boldsymbol{x}(t))$ is a generic function of the vector $\boldsymbol{x}(t) = \{x_i(t)\}$ of all concentrations and determines the drift of x_i. In the

biochemical context it gives the rate of change in x_i due to reactions with other species and includes the relevant reaction rates. A term $-\lambda_i x_i$ has been included that drives each x_i back to zero, with λ_i having the meaning of a decay rate. Finally, $\xi_i(t)$ is Gaussian white noise with the properties

$$\langle \xi_i \rangle = 0 \qquad \langle \xi_i(t)\xi_j(t') \rangle = \Sigma_{ii}\delta_{ij}\delta(t-t') \tag{2.3.2}$$

The Kronecker delta δ_{ij} signifies that each variable x_i has independent noise acting on it. Correlations in the noise could be allowed for by extending the matrix $\Sigma_{ii}\delta_{ij}$ to one having nonzero off-diagonal entries, but become difficult to express in terms of the local parameters that define the core of the extended Plefka expansion, as will be explained below.

After discretizing time in elementary time steps Δ, a dynamical partition function for this system can be written in the Martin–Siggia–Rose–Janssen–De Dominicis (MSRJD) functional integral formalism [19–21]

$$\begin{aligned} Z &= \left\langle \int \prod_{it} dx_i(t)\delta(x_i(t+\Delta) - x_i(t) - \Delta[-\lambda_i x_i(t) + \phi_i(\boldsymbol{x}(t)) + \xi_i(t)]) \right\rangle_{\boldsymbol{\xi}} = \\ &= \left\langle \int \prod_{it} \frac{dx_i(t)d\hat{x}_i(t)}{2\pi} \mathrm{e}^{\mathrm{i}\hat{x}_i(t)(x_i(t+\Delta)-x_i(t)-\Delta[-\lambda_i x_i(t)+\phi_i(\boldsymbol{x}(t))+\xi_i(t)])} \right\rangle_{\boldsymbol{\xi}} \end{aligned} \tag{2.3.3}$$

We use the Itô convention [10] to discretize the noise, where $\xi_i(t)$ above is to be read as the average of the continuous-time noise over the time interval $[t, t+\Delta]$, which has covariance

$$\langle \xi_i(t)\xi_i(t') \rangle = \frac{1}{\Delta}\Sigma_{ii}\delta_{tt'} \tag{2.3.4}$$

Here $\delta_{tt'}/\Delta$ is the discrete-time replacement of $\delta(t-t')$. The Δ term shows that the actual contribution of the covariance $\boldsymbol{\Sigma}$ does not depend on time discretization (i.e. it is $O(1)$ in the $\Delta \to 0$ limit). The average over the white noise can then be performed by applying a standard Gaussian identity

$$\langle \mathrm{e}^{\mathrm{i}\Delta\hat{\boldsymbol{x}}^{\mathrm{T}}\cdot\boldsymbol{\xi}} \rangle_{\boldsymbol{\xi}} = \mathrm{e}^{-\Delta\,\hat{\boldsymbol{x}}^T\boldsymbol{\Sigma}\hat{\boldsymbol{x}}/2} \tag{2.3.5}$$

To develop a Plefka expansion, we now need to consider which averages should be constrained in the relevant Legendre transform. By reinterpreting the static TAP equations from the perspective of a cavity argument [37], one would obtain marginals where the covariance of the cavity field and a quadratic term for the spins is present. These are effectively constant in the case of Ising spins ($s_i^2 = 1$) but should be explicitly taken into account for continuous variables (even in a static problem) and for tracking time dependencies (see [37] for spin dynamics).

Let us now introduce some shorthands to explain in intuitive terms the logic beyond the "extended" Plefka expansion, connecting it to the version for equilibrium systems outlined in section 2.2. We

denote

$$\hat{\boldsymbol{m}} = \{\boldsymbol{x}, -\mathrm{i}\hat{\boldsymbol{x}}, \boldsymbol{x}\boldsymbol{x}, -\mathrm{i}\hat{\boldsymbol{x}}\boldsymbol{x}, \mathrm{i}\hat{\boldsymbol{x}}\mathrm{i}\hat{\boldsymbol{x}}\} \tag{2.3.6a}$$

$$\boldsymbol{m} = \{\boldsymbol{\mu}, -\mathrm{i}\hat{\boldsymbol{\mu}}, \boldsymbol{C}, \boldsymbol{R}, \boldsymbol{B}\} \tag{2.3.6b}$$

$$\boldsymbol{h}_\alpha = \{\boldsymbol{\Psi}_\alpha, \boldsymbol{l}_\alpha, \hat{\boldsymbol{C}}_\alpha, \hat{\boldsymbol{R}}_\alpha, \hat{\boldsymbol{B}}_\alpha\} \tag{2.3.6c}$$

Here $\hat{\boldsymbol{m}}$ is a compact notation for the quantities whose averages we will constrain, consisting of the $x_i(t)$, $\mathrm{i}\hat{x}_i(t)$ and all their products involving the same degree of freedom or "site" i. It is the inclusion of these products that extends our approach beyond the standard applications of the Plefka method, where only first order moments such as magnetizations are constrained. We indicate by $\boldsymbol{m}$ the constrained values of the relevant averages, which are the order parameters of the theory, and by $\boldsymbol{h}_\alpha$ the conjugate fields. $\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}, \boldsymbol{C}, \boldsymbol{R}, \boldsymbol{B}$ summarize the various groups of order parameters defined as follows

$$\mu_i(t) = \langle x_i(t)\rangle_\alpha \tag{2.3.7a}$$

$$\hat{\mu}_i(t) = \langle \hat{x}_i(t)\rangle_\alpha \tag{2.3.7b}$$

$$C_i(t,t') = \langle x_i(t)x_i(t')\rangle_\alpha \tag{2.3.7c}$$

$$R_i(t',t) = -\mathrm{i}\langle \hat{x}_i(t)x_i(t')\rangle_\alpha \tag{2.3.7d}$$

$$B_i(t,t') = -\langle \hat{x}_i(t)\hat{x}_i(t')\rangle_\alpha \tag{2.3.7e}$$

We denote the corresponding groups of conjugate fields by $\boldsymbol{\Psi}_\alpha, \boldsymbol{l}_\alpha, \hat{\boldsymbol{C}}_\alpha, \hat{\boldsymbol{R}}_\alpha, \hat{\boldsymbol{B}}_\alpha$. All exhibit a strictly local dependence on the species (i.e. all diagonal).

The second order quantities we are constraining involve firstly the (disconnected, local) two-time correlation functions $C_i(t,t')$. From general results for MSRJD path integrals [38] it follows that $R_i(t',t)$ has the meaning of a local response of $x_i(t')$ to a perturbing field $-\mathrm{i}\hat{x}_i(t)$ applied at some earlier time; it should therefore be non-vanishing only for $t' > t$. $B_i(t,t')$, finally, is expected to vanish for all times t and t', as is $\hat{\mu}_i(t)$; both follow from the fact that the dynamical partition function remains equal to unity when generating terms linear in $\hat{x}_i(t)$ are added in the exponent (we refer to [38] for a derivation from the normalization condition).

To define the Plefka free energy, note that after the noise average has been carried out, our partition function can be written in the form $Z = \int D\boldsymbol{x}D\hat{\boldsymbol{x}}\, \mathrm{e}^{\mathcal{H}_\alpha}$ with a suitable Hamiltonian (or action) $\mathcal{H}_\alpha$ for the stochastic dynamics. Here $D\boldsymbol{x}D\hat{\boldsymbol{x}}$ is a shorthand for the integral $\prod_{it} \frac{dx_i(t)d\hat{x}_i(t)}{2\pi}$ and corresponds to the trace over spins. As in the equilibrium calculation one now defines the Plefka energy G_α as

$$G_\alpha(\boldsymbol{m}) = \hat{G}_\alpha(\boldsymbol{m}, \boldsymbol{h}_\alpha(\boldsymbol{m})) = \ln \int D\boldsymbol{x}D\hat{\boldsymbol{x}}\, \mathrm{e}^{\Xi_\alpha} \tag{2.3.8}$$

where

$$\Xi_\alpha = \mathcal{H}_\alpha + \boldsymbol{h}_\alpha \cdot (\hat{\boldsymbol{m}} - \boldsymbol{m}) \tag{2.3.9}$$

Explicitly, one has for our system and with the extended set of Plefka order parameters

$$\begin{aligned}
\Xi_\alpha &= \sum_{it} \mathrm{i}\hat{x}_i(t)\big(x_i(t+\Delta) - x_i(t) + \Delta\lambda_i x_i(t) - \alpha\Delta\phi_i(\boldsymbol{x}(t))\big) + \Delta\sum_{it}\psi_{i\alpha}(t)\big(x_i(t) - \mu_i(t)\big) + \\
&- \Delta\sum_{it} l_{i\alpha}(t)\big(\mathrm{i}\hat{x}_i(t) - \mathrm{i}\hat{\mu}_i(t)\big) + \Delta^2\sum_{itt'}\hat{C}_{i\alpha}(t,t')\big(x_i(t)x_i(t') - C_i(t,t')\big) + \\
&+ \Delta^2\sum_{itt'}\hat{R}_{i\alpha}(t,t')\big(-\mathrm{i}\hat{x}_i(t)x_i(t') - R_i(t',t)\big) + \frac{\Delta^2}{2}\sum_{itt'}\hat{B}_{i\alpha}(t,t')\big(-\hat{x}_i(t)\hat{x}_i(t') - B_i(t,t')\big) + \\
&- \frac{\Delta}{2}\sum_{it}\Sigma_{ii}\hat{x}_i(t)\hat{x}_i(t)
\end{aligned} \tag{2.3.10}$$

where the first and last terms constitute the Hamiltonian $\mathcal{H}_\alpha$. Note that we have inserted powers of Δ in such a way as to keep the fields of order unity in the continuous time limit $\Delta \to 0$. The parameter α characterizes the strength of the interactions as in the equilibrium case, here via ϕ_i; the linear self-interaction via $-\lambda_i x_i$ is tractable and so is left as part of the non-interacting baseline. Our aim will be to use a second-order Plefka expansion to derive an effective non-interacting description of our system, where the interactions between variables are replaced by additional coloured noise and a coupling of each variable to its past.

In analogy with the equilibrium expansion, the fields $\boldsymbol{h}_\alpha$ are determined by extremization of $\hat{G}_\alpha$. Once G_α has been found, the fields can be retrieved from $\boldsymbol{h}_\alpha = -\partial G_\alpha/\partial \boldsymbol{m}$ and order parameters of the original system dynamics can be found from the condition $\boldsymbol{h}_\alpha = 0$. Split into the various order parameter groups, the derivatives of G_α read

$$\psi_{i\alpha}(t) = -\frac{1}{\Delta}\frac{\partial G_\alpha}{\partial \mu_i(t)} \tag{2.3.11a}$$

$$-\mathrm{i}l_{i\alpha}(t) = -\frac{1}{\Delta}\frac{\partial G_\alpha}{\partial(\hat{\mu}_i(t))} \tag{2.3.11b}$$

$$\hat{R}_{i\alpha}(t,t') = -\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial R_i(t',t)} \tag{2.3.11c}$$

$$\hat{C}_{i\alpha}(t,t') = -\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial C_i(t,t')} \tag{2.3.11d}$$

$$\hat{B}_{i\alpha}(t,t') = -\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial B_i(t,t')} \tag{2.3.11e}$$

We now proceed with the Plefka expansion of G_α around $\alpha = 0$ up to second order, and define a set of effective fields $\boldsymbol{h}^{\mathrm{eff}}$ as in (2.2.14). These provide the effective non-interacting description of the true interacting dynamics, whereby with these fields at $\alpha = 0$ the order parameters have the same values as in the interacting system. As the $\boldsymbol{h}^{\mathrm{eff}}$ themselves depend on the order parameters, this typically leads to nonlinear self-consistency equations, which are the analogues of the MF and TAP equations for the SK model.

The above makes clear why we have introduced only fields depending on a single site: this assumption guarantees that the effective dynamics will be non-interacting. We also see now why correlations between the noises ξ_i affecting the different x_i would complicate matters: the correlations $C_{ij}(t,t')$ would be non-local even at $\alpha = 0$, and determined only in a very indirect way from the local order parameters $C_i(t,t')$. In the application to biochemical reaction networks there generally are non-trivial noise correlations as discussed in section 2.6 below, and further work would be required to understand how best to deal with those.

2.3.1 Structure of the non-interacting problem

In the logic explained above, the intractable part of the interactions becomes condensed into local fields that describe the effective single-site dynamics. These effective fields $\psi_i^{\text{eff}}(t)$, $l_i^{\text{eff}}(t)$, $\hat{R}_i^{\text{eff}}(t,t')$, $\hat{C}_i^{\text{eff}}(t,t')$, $\hat{B}_i^{\text{eff}}(t,t')$ appear in the corresponding effective action Ξ^{eff}

$$\begin{aligned}\Xi^{\text{eff}} &= \sum_{it} \mathrm{i}\hat{x}_i(t)\Big(x_i(t+\Delta) - x_i(t) + \Delta\lambda_i x_i(t)\Big) - \frac{\Delta}{2}\sum_{it}\Sigma_{ii}\hat{x}_i(t)\hat{x}_i(t) + \Delta\sum_{it}\psi_i^{\text{eff}}(t)\Big(x_i(t) - \mu_i(t)\Big) + \\ &- \Delta\sum_{it} l_i^{\text{eff}}(t)\Big(\mathrm{i}\hat{x}_i(t) - \mathrm{i}\hat{\mu}_i(t)\Big) + \Delta^2\sum_{itt'}\hat{C}_i^{\text{eff}}(t,t')\Big(x_i(t)x_i(t') - C_i(t,t')\Big) + \\ &+ \Delta^2\sum_{itt'}\hat{R}_i^{\text{eff}}(t,t')\Big(-\mathrm{i}\hat{x}_i(t)x_i(t') - R_i(t',t)\Big) + \frac{\Delta^2}{2}\sum_{itt'}\hat{B}_i^{\text{eff}}(t,t')\Big(-\hat{x}_i(t)\hat{x}_i(t') - B_i(t,t')\Big)\end{aligned} \tag{2.3.12}$$

To get the generic self-consistency equations for our order parameters, we should in principle evaluate the averages $\mu_i(t)$, $\hat{\mu}_i(t)$, $C_i(t,t')$, $R_i(t',t)$ and $B_i(t,t')$ for this action. The result is the analogue of what for an equilibrium spin problem is $m_i = \tanh(h_i^{\text{ext}} + h_i^{\text{eff}})$.

To simplify this procedure, one can make the natural (see above) assumptions that the solution of the self-consistency equations will obey $\hat{\mu}_i(t) = 0$, $B_i(t,t') = 0$ and $R_i(t,t') = 0$ for $t' \geq t$; the vanishing of the response at equal times is a generic consequence of the Itô discretization. We will have to check that these assumptions are self-consistent. As we show below, they imply $\psi_i^{\text{eff}}(t) = 0$, $\hat{C}_i^{\text{eff}}(t,t') = 0$ and $\hat{R}_i^{\text{eff}}(t,t') = 0$ for $t' \geq t$ so that the effective action reduces to

$$\begin{aligned}\Xi^{\text{eff}} &= \sum_{it}\mathrm{i}\hat{x}_i(t)\Big[x_i(t+\Delta) - x_i(t) + \Delta\Big(\lambda_i x_i(t) - l_i^{\text{eff}}(t) - \Delta\sum_{t'<t}\hat{R}_i^{\text{eff}}(t,t')x_i(t')\Big)\Big] + \\ &- \frac{\Delta^2}{2}\sum_{itt'}\hat{B}_i^{\text{eff}}(t,t')\hat{x}_i(t)\hat{x}_i(t') - \frac{\Delta}{2}\sum_{it}\Sigma_{ii}\hat{x}_i(t)\hat{x}_i(t)\end{aligned} \tag{2.3.13}$$

This is exactly the action for the Langevin dynamics

$$\frac{x_i(t+\Delta) - x_i(t)}{\Delta} = -\lambda_i x_i(t) + l_i^{\text{eff}}(t) + \Delta\sum_{t'}\hat{R}_i^{\text{eff}}(t,t')x_i(t') + \xi_i(t) + \chi_i(t) \tag{2.3.14}$$

where $\boldsymbol{\chi}$ is a coloured, local Gaussian noise with

$$\langle\chi_i\rangle = 0 \qquad \langle\chi_i(t)\chi_i(t')\rangle = \hat{B}_i^{\text{eff}}(t,t') \tag{2.3.15}$$

Note that the covariance of this effective noise is defined exactly so that the quadratic terms in $\hat{x}_i(t)$ in Ξ^{eff} arise from averaging over χ_i

$$e^{-\Delta^2 \sum_{tt'} \hat{x}_i(t)\hat{B}_i^{\text{eff}}(t,t')\hat{x}_i(t')/2} = \langle e^{i\Delta \sum_t \hat{x}_i(t)\chi_i(t)} \rangle_\chi \tag{2.3.16}$$

The remainder of the analysis is easier to carry out in the continuous time-limit $\Delta \to 0$. The effective equation of motion becomes

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + I_i^{\text{eff}}(t) + \int_0^t dt' \hat{R}_i^{\text{eff}}(t,t')x_i(t') + \xi_i(t) + \chi_i(t) \tag{2.3.17}$$

which shows that $\hat{R}_i^{\text{eff}}(t,t')$ plays the role of a memory function. Because this dynamics is causal, it does indeed give $\hat{\mu}_i = 0$, $B_i(t,t') = 0$ and $R_i(t,t') = 0$ for $t' \geq t$, and so our original assumptions about the order parameter values are self-consistent.

It remains to obtain the equations for the nonzero order parameters $\mu_i(t)$, $R_i(t,t')$ for $t > t'$, and $C_i(t,t')$. For the means we have by simple averaging over the zero mean noises ξ_i and χ_i

$$\frac{d\mu_i(t)}{dt} = -\lambda_i \mu_i(t) + \int_0^t dt' \hat{R}_i^{\text{eff}}(t,t')\mu_i(t') + I_i^{\text{eff}}(t) \tag{2.3.18}$$

For the responses, standard results for linear dynamics with Gaussian noise give

$$\frac{\partial R_i(t,t')}{\partial t} = \frac{\partial \dot{\mu}_i(t)}{\partial I_i^{\text{eff}}(t')} = -\lambda_i R_i(t,t') + \int_{t'}^t dt'' \hat{R}_i^{\text{eff}}(t,t'')R_i(t'',t') + \delta(t-t') \tag{2.3.19}$$

For the correlations it makes sense to consider the connected version $\delta C_i(t',t) = C_i(t',t) - \mu_i(t')\mu_i(t)$, which obeys

$$\frac{\partial \delta C_i(t,t')}{\partial t} = -\lambda_i \delta C_i(t,t') + \int_{t'}^t dt'' \hat{R}_i^{\text{eff}}(t,t'')\delta C_i(t'',t') + \int_0^{t'} dt'' R_i(t',t'')(\hat{B}_i^{\text{eff}}(t,t'') + \Sigma_{ii}\delta(t-t'')) \tag{2.3.20}$$

These order parameters $\mu_i(t)$, $\delta C_i(t,t')$ and $R_i(t',t)$ are uniquely determined from the above equations when supplemented with initial values $\mu_i(0)$ and $\delta C_i(0,0)$, which we assume are given as part of the specification of our system. Let us look at this derivation in slight more detail. For the response, it is useful to notice that differentiation w.r.t. a fluctuating external field is equivalent to differentiation w.r.t. the noise

$$R_i(t,t') = \frac{\partial \mu_i(t)}{\partial I_i^{\text{eff}}(t')} = \left\langle \frac{\partial x_i(t)}{\partial(\xi+\chi)_i(t')} \right\rangle \tag{2.3.21}$$

We can thus derive the dynamics of response functions

$$\begin{aligned} \frac{\partial R_i(t,t')}{\partial t} = \frac{\partial}{\partial t}\left\langle \frac{\partial x_i(t)}{\partial(\xi+\chi)_i(t')} \right\rangle = \left\langle \frac{\partial \dot{x}_i(t)}{\partial(\xi+\chi)_i(t')} \right\rangle = \\ -\lambda_i R_i(t,t') + \int_{t'}^t dt'' \hat{R}_i^{\text{eff}}(t,t'')R_i(t'',t') + \delta(t-t') \end{aligned} \tag{2.3.22}$$

For connected correlations, we first write down explicitly the effective equation for $\delta x_i(t) = x_i(t) - \mu_i(t)$

$$\frac{d\,\delta x_i(t)}{dt} = -\lambda_i \delta x_i(t) + \int_0^t dt' \hat{R}_i^{\text{eff}}(t,t')\delta x_i(t') + \xi_i(t) + \chi_i(t) \tag{2.3.23}$$

Then we multiply by $\delta x_i(t')$ and average to get (2.3.20). One uses a theorem for Gaussian variables

$$\begin{aligned}\langle \delta x_i(t')(\xi+\chi)_i(t)\rangle &= \int_0^{t'} dt'' \Big\langle \frac{\partial x_i(t')}{\partial (\xi+\chi)_i(t'')} \Big\rangle \langle (\xi+\chi)_i(t'')(\xi+\chi)_i(t)\rangle = \\ &= \int_0^{t'} dt'' R_i(t',t'')(\hat{B}_i^{\text{eff}}(t,t'') + \Sigma_{ii}\delta(t-t''))\end{aligned} \tag{2.3.24}$$

A proof of this theorem is provided in [39].

2.3.2 First order: Mean Field equations

As explained above the equilibrium case, see equations (2.2.16) and (2.2.17), the first order correction in α to the Plefka free energy is

$$G^1 = \frac{\partial G_\alpha}{\partial \alpha}\Big|_{\alpha=0} = \Big\langle \frac{d\Xi_\alpha}{d\alpha} \Big\rangle_0 \tag{2.3.25}$$

or explicitly

$$\begin{aligned} G^1 = &-\Delta \sum_{it} \langle \mathrm{i}\hat{x}_i(t)\phi_i(\boldsymbol{x}(t))\rangle_\alpha + \Delta \sum_{it} \frac{\partial \psi_{i\alpha}(t)}{\partial \alpha} \langle x_i(t) - \mu_i(t)\rangle_\alpha - \Delta \sum_{it} \frac{\partial l_{i\alpha}(t)}{\partial \alpha} \langle \mathrm{i}\hat{x}_i(t) - \mathrm{i}\hat{\mu}_i(t)\rangle_\alpha + \\ &+ \Delta^2 \sum_{itt'} \frac{\partial \hat{C}_{i\alpha}(t,t')}{\partial \alpha} \langle x_i(t)x_i(t') - C_i(t,t')\rangle_\alpha + \Delta^2 \sum_{itt'} \frac{\partial \hat{R}_{i\alpha}(t,t')}{\partial \alpha} \langle -\mathrm{i}\hat{x}_i(t)x_i(t') - R_i(t',t)\rangle_\alpha + \\ &+ \frac{\Delta^2}{2} \sum_{itt'} \frac{\partial \hat{B}_i(t,t')}{\partial \alpha} \langle -\hat{x}_i(t)\hat{x}_i(t') - B_i(t,t')\rangle_\alpha \Big|_{\alpha=0} \end{aligned} \tag{2.3.26}$$

where all the factors multiplying the derivatives of the fields vanish by definition (2.3.7).

G^1 thus reduces to

$$G^1 = -\Delta \sum_{it} \langle \mathrm{i}\hat{x}_i(t)\phi_i(\boldsymbol{x}(t))\rangle_0 \tag{2.3.27}$$

For the sake of brevity we drop the subscript 0: all averages below are to be taken at $\alpha = 0$ unless otherwise specified. To find G^1 explicitly, consider first a generic vector $\boldsymbol{z} = \{z_a\}$ of Gaussian variables with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}$. Then by integration by parts

$$\langle \delta z_a \phi(\boldsymbol{z})\rangle = \sum_b \Gamma_{ab} \langle \partial_{z_b}\phi(\boldsymbol{z})\rangle \tag{2.3.28}$$

where $\delta z_a = z_a - \mu_a$. Applying this first identity to our case gives

$$\langle \delta \hat{x}_i(t)\phi_i(\boldsymbol{x}(t))\rangle = \mathrm{i}\delta R_i(t,t) \Big\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \Big\rangle \tag{2.3.29}$$

where $\delta\hat{x}_i = \hat{x}_i - \hat{\mu}_i$ and $\delta R_i(t,t') = R_i(t,t') + \mathrm{i}\hat{\mu}_i(t')\mu_i(t))$ is the connected response function. As a consequence,

$$G^1 = -\Delta \sum_{it} \left(\mathrm{i}\hat{\mu}_i(t)\langle\phi_i(\boldsymbol{x}(t))\rangle - \delta R_i(t,t)\left\langle \frac{\partial\phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \right\rangle \right) \tag{2.3.30}$$

While not fully explicit, the value of this expression is fully determined by our order parameters; specifically the averages over $\boldsymbol{x}(t)$ are over independent Gaussian variables $x_i(t)$ with mean $\mu_i(t)$ and variance $\delta C_i(t,t) = C_i(t,t) - \mu_i^2(t)$.

We can now obtain the first order (in α) conjugate fields, which are the negative derivatives of G^1 w.r.t. the order parameters

$$\boldsymbol{h}^1 = -\frac{1}{\Delta^n}\frac{\partial G^1}{\partial \boldsymbol{m}}, \qquad \boldsymbol{h}^1 = \{\psi_i^1(t), l_i^1(t), \hat{C}_i^1(t,t'), \hat{R}_i^1(t',t), \hat{B}_i^1(t,t')\} \tag{2.3.31}$$

where according to our convention in the construction of Ξ_α, the exponent $n = 1$ for linear order parameters and $n = 2$ for quadratic ones. Explicitly we obtain

$$\begin{aligned}
\psi_i^1(t) = \quad & \sum_j \left(\mathrm{i}\hat{\mu}_j(t)\frac{\partial\langle\phi_j(\boldsymbol{x}(t))\rangle}{\partial\mu_i(t)} - \delta R_j(t,t)\frac{\partial}{\partial\mu_i(t)}\left\langle\frac{\partial\phi_j(\boldsymbol{x}(t))}{\partial x_j(t)}\right\rangle\right) + \\
& - \mathrm{i}\hat{\mu}_i(t)\left\langle\frac{\partial\phi_i(\boldsymbol{x}(t))}{\partial x_i(t)}\right\rangle
\end{aligned} \tag{2.3.32a}$$

$$l_i^1(t) = \quad \mu_i(t)\left\langle\frac{\partial\phi_i(\boldsymbol{x}(t))}{\partial x_i(t)}\right\rangle - \langle\phi_i(\boldsymbol{x}(t))\rangle \tag{2.3.32b}$$

$$\hat{C}_i^1(t,t') = \quad \frac{1}{\Delta}\sum_j \left(\mathrm{i}\hat{\mu}_j(t)\frac{\partial\langle\phi_j(\boldsymbol{x}(t))\rangle}{\partial C_i(t,t)} - \delta R_j(t,t)\frac{\partial}{\partial C_i(t,t)}\left\langle\frac{\partial\phi_j(\boldsymbol{x}(t))}{\partial x_j(t)}\right\rangle\right)\delta_{tt'} \tag{2.3.32c}$$

$$\hat{R}_i^1(t,t') = -\frac{1}{\Delta}\left\langle\frac{\partial\phi_i(\boldsymbol{x}(t))}{\partial x_i(t)}\right\rangle\delta_{tt'} \tag{2.3.32d}$$

$$\hat{B}_i^1(t,t') = \quad 0 \tag{2.3.32e}$$

Using the general identity for Gaussian variables $\boldsymbol{z} = \{z_a\}$ with means μ_a

$$\partial_{\mu_a}\langle\phi(\boldsymbol{z})\rangle = \langle\partial_{z_a}\phi(\boldsymbol{z})\rangle \tag{2.3.33}$$

the first average in the expression for ψ_i^1 could also be written as $\langle\partial\phi_j(\boldsymbol{x}(t))/\partial x_i(t)\rangle$. The effective fields defining the effective non-interacting dynamics are now $\boldsymbol{h}^{\text{eff}} = -\alpha\boldsymbol{h}^1$. To evaluate these we can exploit that the final order parameter values should obey $\hat{\mu}_i(t) = 0$ and $R_i(t,t) = 0$, hence also $\delta R_i(t,t) = 0$. This then gives $\psi_i^1(t) = 0$ and $\hat{C}_i^1(t,t) = 0$ so that also the corresponding effective fields vanish, as anticipated above in our general discussion of the effective non-interacting dynamics. Note that it is important to make the above simplifying assumptions only in the final expressions for the effective fields, not already in G^1 as derivatives w.r.t. e.g. $\hat{\mu}_i(t)$ do contribute to the effective fields.

The only remaining nonzero effective fields at this stage are $l_i^{\text{eff}}(t) = -\alpha l_i^1(t)$ and $\hat{R}_i^{\text{eff}}(t,t') = -\alpha \hat{R}_i^1(t,t')$. We insert these into (2.3.13) to get the mean field equations for the now effectively non-interacting degrees of freedom $x_i(t)$

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \alpha\Big\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \Big\rangle (x_i(t) - \mu_i(t)) + \alpha\langle \phi_i(\boldsymbol{x}(t))\rangle + \xi_i(t) \tag{2.3.34}$$

Not unexpectedly for an effective linear dynamics, the interaction term $\phi_i(\boldsymbol{x}(t))$ has here effectively been linearized in deviations of $x_i(t)$ from its mean. The self-consistency equation for this mean reads

$$\frac{d\mu_i(t)}{dt} = -\lambda_i \mu_i(t) + \alpha\langle \phi_i(\boldsymbol{x}(t))\rangle \tag{2.3.35}$$

The equations for the equal-time correlations $C_i(t,t)$ can be obtained from the equation of motion for the fluctuations around the mean $\delta x_i(t) = x_i(t) - \mu_i(t)$

$$\frac{d\delta x_i(t)}{dt} = \Big(-\lambda_i + \alpha\Big\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \Big\rangle \Big) \delta x_i(t) + \xi_i(t) \tag{2.3.36}$$

This gives directly, in the standard manner for an Ornstein-Uhlenbeck process with time-dependent drift,

$$\frac{d\delta C_i(t,t)}{dt} = 2\Big(-\lambda_i + \alpha\Big\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \Big\rangle \Big) \delta C_i(t,t) + \Sigma_{ii} \tag{2.3.37}$$

In general, the above equations need to be solved jointly for the $2N$ time-dependent order parameters $\mu_i(t)$ and $C_i(t,t)$; this is because the average of $\partial\phi_i/\partial x_i$ generically depends on both means and variances. The case of purely linear interactions, where $\phi_i = \sum_{j\neq i} K_{ij} x_j$, is an obvious exception: here the equations for the means do not involve the variances so can be solved separately.

It is worth commenting at this stage how our first order result compares with that of a conventional Plefka approach that constrains only the first moments $\mu_i(t)$ and $\hat{\mu}_i(t)$. The effective field terms in the effective dynamical action are then linear in $x_i(t)$ and $\hat{x}_i(t)$. This means that all second order fluctuation statistics remain as in a non-interacting problem. In particular, $\delta C_i(t,t')$ and $R_i(t,t')$ do not feel any effect of the non-trivial drift ϕ_i. The second term in the brackets in the r.h.s. of (2.3.36) would be absent, and the interaction term ϕ_i would only appear via its average. Already to first order in α it is clear, then, that the extended Plefka approach captures qualitatively more of the dynamics of the interacting system than a conventional Plefka method constraining linear averages.

2.3.3 Second order: TAP equations

The second order of the Plefka free energy can be evaluated starting from the equality (2.2.19)

$$G^2 = \frac{\partial^2 G_\alpha}{\partial \alpha^2}\Big|_{\alpha=0} = \Big\langle \Big(\delta \frac{d\Xi_\alpha}{d\alpha} \Big)^2 \Big\rangle_0 \tag{2.3.38}$$

Including the first order fields in the effective action, with the prefactor α, and taking $\frac{d\Xi_\alpha}{d\alpha}$ at $\alpha = 0$ gives

$$\frac{d\Xi_\alpha}{d\alpha}\Big|_{\alpha=0} = -\Delta \sum_{it} \mathrm{i}\hat{x}_i(t)\phi_i(\boldsymbol{x}(t)) + \Delta \sum_{it} \psi_i^1(t)\Big(x_i(t) - \mu_i(t)\Big) - \Delta \sum_{it} l_i^1(t)\Big(\mathrm{i}\hat{x}_i(t) - \mathrm{i}\hat{\mu}_i(t)\Big) +$$
$$+\Delta^2 \sum_{itt'} \hat{C}_i^1(t,t')\Big(x_i(t)x_i(t') - C_i(t,t')\Big) + \Delta^2 \sum_{itt'} \hat{R}_i^1(t,t')\Big(-\mathrm{i}\hat{x}_i(t)x_i(t') - R_i(t',t)\Big) \tag{2.3.39}$$

While the following analysis can be carried out for general drift $\phi_i(\boldsymbol{x})$, we will restrict the scenario slightly by assuming that

$$\frac{\partial \phi_i(\boldsymbol{x})}{\partial x_i} = 0 \tag{2.3.40}$$

as this significantly reduces the number of terms in the expressions (see appendix C for full derivation without this simplification). Intuitively, we are assuming that $\phi_i(\boldsymbol{x})$ is a function only of the other variables x_j; equivalently, x_i interacts with itself only via the linear term $-\lambda_i x_i$. In the later steps of the calculation, from (2.3.43), we will add the assumption that the drift ϕ_i is an additive combinations of functions of the other variables x_j, i.e. of the form $\phi_i(\boldsymbol{x}) = \sum_{j\neq i} g_{ij}(x_j)$. The above expressions for the first order conjugate fields then simplify to

$$\psi_i^1(t) = \sum_j \mathrm{i}\hat{\mu}_j(t) \frac{\partial \langle \phi_j(\boldsymbol{x}(t))\rangle}{\partial \mu_i(t)} \tag{2.3.41a}$$

$$l_i^1(t) = -\langle \phi_i(\boldsymbol{x}(t))\rangle \tag{2.3.41b}$$

$$\hat{C}_i^1(t,t') = \frac{1}{\Delta}\sum_j \mathrm{i}\hat{\mu}_j(t) \frac{\partial \langle \phi_j(\boldsymbol{x}(t))\rangle}{\partial C_i(t,t')}\delta_{tt'} \tag{2.3.41c}$$

$$\hat{R}_i^1(t,t') = 0 \tag{2.3.41d}$$

$$\hat{B}_i^1(t,t') = 0 \tag{2.3.41e}$$

Inserting these into (2.3.39) one finds

$$\delta\frac{d\Xi_\alpha}{d\alpha} = \frac{d\Xi_\alpha}{d\alpha} - \left\langle \frac{d\Xi_\alpha}{d\alpha}\right\rangle_0 = -\Delta \sum_{it}\Big[\mathrm{i}\delta\hat{x}_i(t)\delta\phi_i(\boldsymbol{x}(t)) + \mathrm{i}\hat{\mu}_i(t)\sum_j \Big(\delta\phi_i(\boldsymbol{x}(t)) \tag{2.3.42}$$
$$- \frac{\partial\langle\phi_i(\boldsymbol{x}(t))\rangle}{\partial\mu_j(t)}\delta x_j(t) - \frac{\partial\langle\phi_i(\boldsymbol{x}(t))\rangle}{\partial C_j(t,t)}\delta(x_j(t)x_j(t))\Big)\Big]$$

where $\delta\hat{x}_i(t) = \hat{x}_i(t) - \hat{\mu}_i(t)$ as before and $\delta\phi_i(\boldsymbol{x}(t)) = \phi_i(\boldsymbol{x}(t)) - \langle\phi_i(\boldsymbol{x}(t))\rangle$, while $\delta(x_j(t)x_j(t)) = x_j^2(t) - C_j(t,t)$. To calculate G^2 one now needs to square this and evaluate the relevant averages, expressing them in terms of the relevant order parameters (2.3.7). The square will give correlations of terms at different sites i and j and different times t and t'. Because the averages are taken at $\alpha = 0$, there are no correlations between variables at different sites i. For the same reason all

statistics are Gaussian, and one can use Wick's theorem to reduce all higher order moments to first and second order ones. Once G^2 (see appendix) has been found, the $O(\alpha^2)$ corrections for the fields can be calculated from

$$\boldsymbol{h}^2 = -\frac{1}{\Delta^n}\frac{\partial G^2}{\partial \boldsymbol{m}} \tag{2.3.43}$$

which is just the second order analogue of (2.3.31). With these general expressions for the fields we obtained, one can again impose the physical constraints on the order parameters, i.e. $\hat{\mu}_i(t) = 0$, $\delta R_i(t,t') = 0$ for $t' \geq t$ and $\delta B_i(t,t') = B_i(t,t') + \hat{\mu}_i(t)\hat{\mu}_i(t') = 0$. We omit the details and write directly the final simplified form of the second order fields

$$\psi_i^2(t) = 0 \tag{2.3.44a}$$

$$l_i^2(t) = 2\Delta \sum_{jt'} \left\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \frac{\partial \phi_j(\boldsymbol{x}(t'))}{\partial x_i(t')} \right\rangle \mu_i(t') R_j(t,t') \tag{2.3.44b}$$

$$\hat{R}_i^2(t,t') = -2 \sum_{j} \left\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \frac{\partial \phi_j(\boldsymbol{x}(t'))}{\partial x_i(t')} \right\rangle R_j(t,t') \tag{2.3.44c}$$

$$\hat{C}_i^2(t,t') = 0 \tag{2.3.44d}$$

$$\hat{B}_i^2(t,t') = -\langle \delta\phi_i(\boldsymbol{x}(t)) \delta\phi_i(\boldsymbol{x}(t')) \rangle \tag{2.3.44e}$$

These fields, multiplied by $-\frac{\alpha^2}{2}$, give the second order contributions to the effective fields in the non-interacting dynamical action, $\boldsymbol{h}^{\text{eff}} = -\alpha \boldsymbol{h}^1 - \frac{\alpha^2}{2}\boldsymbol{h}^2$. One sees that $\psi_i^{\text{eff}}(t)$ and $\hat{C}_i^{\text{eff}}(t,t')$ remain identically zero also to second order, while the nonzero effective fields are, in the continuous time limit $\Delta \to 0$

$$l_i^{\text{eff}}(t) = \alpha \langle \phi_i(\boldsymbol{x}(t)) \rangle - \alpha^2 \int_0^t dt' \sum_j \left\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \frac{\partial \phi_j(\boldsymbol{x}(t'))}{\partial x_i(t')} \right\rangle \mu_i(t') R_j(t,t') \tag{2.3.45}$$

$$\hat{R}_i^{\text{eff}}(t,t') = \alpha^2 \sum_j \left\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \frac{\partial \phi_j(\boldsymbol{x}(t'))}{\partial x_i(t')} \right\rangle R_j(t,t') \tag{2.3.46}$$

$$\hat{B}_i^{\text{eff}}(t,t') = \alpha^2 \langle \delta\phi_i(\boldsymbol{x}(t)) \delta\phi_i(\boldsymbol{x}(t')) \rangle \tag{2.3.47}$$

We notice that the causality structure of $\hat{R}_i^{\text{eff}}(t,t')$ is directly related to that of $R_i(t,t')$, i.e. both are nonzero only when the second time argument is smaller than the first. (In the first order calculation we had in addition found a nonzero equal-time value for $\hat{R}_i^{\text{eff}}(t,t')$ but this was due to a self-interaction that we have since assumed to be zero.) Substituting the fields into Ξ^{eff} (2.3.13), we obtain the uncoupled description of the dynamics to second order in α

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \alpha \langle \phi_i(\boldsymbol{x}(t)) \rangle + \alpha^2 \sum_j \int_0^t dt' \left\langle \frac{\partial \phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \frac{\partial \phi_j(\boldsymbol{x}(t'))}{\partial x_i(t')} \right\rangle R_j(t,t') \delta x_i(t') + \xi_i(t) + \chi_i(t) \tag{2.3.48}$$

The dynamical TAP equations are then the self-consistency equations for the $\mu_i(t)$, $R_i(t,t')$ and $C_i(t,t')$ that result. These are written in their general form in (2.3.18) to (2.3.20) above. What is remarkable is that the integral over the past in (2.3.48) does not contribute to the evolution equation for the means, which as to first order is given by

$$\frac{d\mu_i(t)}{dt} = -\lambda_i\mu_i(t) + \alpha\langle\phi_i(\boldsymbol{x}(t))\rangle \tag{2.3.49}$$

This does not mean, of course, that the actual time courses $\mu_i(t)$ wil be the same in the MF and TAP equations: the TAP equations for the variances $C_i(t,t)$ are different from MF, and these variances affect the average $\langle\phi_i\rangle$ in the evolution of the $\mu_i(t)$.

2.3.4 Linear case

It is instructive to consider this framework for a simple case, i.e. a differential equation with linear couplings

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \sum_j K_{ij}x_j(t) + \xi_i(t) \tag{2.3.50}$$

This corresponds to the choice $\phi_i(\boldsymbol{x}) = \sum_j K_{ij}x_j$ for the drift. We assume throughout that $K_{ii} = 0$, so that there is no self-interaction in ϕ_i.

First order: Mean Field

The first order in α of the Plefka free energy G simplifies from (2.3.27) to

$$G^1 = -\Delta\sum_{it}\mathrm{i}\hat{\mu}_i(t)\sum_j K_{ij}\mu_j(t) \tag{2.3.51}$$

and gives the first order fields

$$\psi_i^1(t) = \sum_j \mathrm{i}\hat{\mu}_j(t)K_{ji} \tag{2.3.52a}$$

$$l_i^1(t) = -\sum_j K_{ij}\mu_j(t) \tag{2.3.52b}$$

$$\hat{R}_i^1(t,t') = \hat{C}_i^1(t,t') = \hat{B}_i^1(t,t') = 0 \tag{2.3.52c}$$

The effective dynamical equation becomes

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \alpha\sum_j K_{ij}\mu_j(t) + \xi_i(t) \tag{2.3.53}$$

and gives for the means the equations of motion

$$\frac{d\mu_i(t)}{dt} = -\lambda_i\mu_i(t) + \alpha\sum_j K_{ij}\mu_j(t) \tag{2.3.54}$$

For $\alpha = 1$ these agree with the exact equations. The second order fluctuation statistics, on the other hand, are unchanged from the non-interacting system at this level of approximation.

Second order: TAP

The effective dynamics to second order in α become, as a special case of (2.3.48)

$$\frac{dx_i(t)}{dt} = -\lambda_i x_i(t) + \alpha \sum_j K_{ij}\mu_j(t) + \alpha^2 \int_0^t dt' \sum_j K_{ij} R_j(t,t') K_{ji} \delta x_i(t') + \xi_i(t) + \chi_i(t) \tag{2.3.55}$$

In the integral term we have arranged the factors to allow a simple intuitive interpretation: a fluctuation δx_i at time t' acts via K_{ji} as an effective field on x_j; at time t this produces a response in x_j modulated by $R_j(t,t')$, which then acts back on x_i via K_{ij}.

Putting $\alpha = 1$, the mean dynamics is identical to the (already exact) MF description

$$\frac{d\mu_i(t)}{dt} = -\lambda_i \mu_i(t) + \sum_j K_{ij}\mu_j(t) \tag{2.3.56}$$

Responses have their temporal evolution governed by (2.3.19)

$$\frac{\partial R_i(t,t')}{\partial t} = -\lambda_i R_i(t,t') + \sum_j \int_{t'}^t dt'' K_{ij} R_j(t,t'') K_{ji} R_i(t'',t') + \delta(t-t') \tag{2.3.57}$$

while for the connected correlations one has, from (2.3.20)

$$\begin{aligned} \frac{\partial\, \delta C_i(t,t')}{\partial t} = & - \lambda_i \delta C_i(t,t') + \sum_j \int_0^t dt'' K_{ij} R_j(t,t'') K_{ji} \delta C_i(t'',t') + \\ & + \Sigma_{ii} R_i(t',t) + \sum_j \int_0^{t'} dt'' R_i(t',t'') K_{ij}^2 \delta C_j(t,t'') \end{aligned} \tag{2.3.58}$$

The last term involves the covariance of the coloured noise $\chi_i(t)$, which is $\sum_j K_{ij}^2 \delta C_j(t,t')$.

2.4 Exactness in the thermodynamic limit

2.4.1 Motivation and setup

The extended Plefka expansion derived above is, of course, an approximation in general because we have truncated the power series expansion in the interaction strength α at second order. We would expect the approximation to become exact, however, provided that the interactions between variables are suitably long-ranged and we take the thermodynamic limit $N \to \infty$ of a large system: a central limit theorem argument then suggests that the interactions have Gaussian statistics as the extended Plefka expansion predicts. The purpose of this section is to study in detail one example of

a model in this class, namely the linear interaction model introduced in section 2.3.4 with random couplings K_{ij}. There are rather more general scenarios where we expect our method to give the exact results, as discussed in section 2.6 below.

We already know (see 2.3.18) that the extended Plefka equations for the means are exact, and will show that the responses and correlations are also predicted correctly by the extended Plefka approach. The exact solution that we work out as our baseline has close similarities with the analysis of the $p = 2$-spin spherical model; see [40] for a detailed study of the latter.

We will focus on the long-time limit $t \to \infty$, where the analysis simplifies because two-time correlations and responses become time translation invariant (TTI), i.e. depend only on time differences. The derivation of the extended Plefka expansion does not of course rely on TTI, and we would expect that the agreement with the exact solution can be demonstrated also for transient relaxation to the steady state.

To be specific, we consider the linear dynamics (2.3.50); this corresponds to the Langevin dynamics of a $p = 2$-spin spherical model where the spins are replaced by arbitrary degrees of freedom $x_i(t)$ interacting in pairs. For the sake of simplicity we assume $\lambda_i = \lambda$ and $\Sigma_{ii} = \Sigma$ for all $i = 1, \ldots, N$, i.e. we take the self-interaction and noise strength as the same for all degrees of freedom. The self-interaction plays the role of the Lagrange multiplier enforcing the spherical constraint in the $p = 2$-spin spherical model: note that in our case it is not time dependent, however, but simply a constant.

We want to proceed with as few restrictive assumptions on the couplings K_{ij} as possible; in fact, nothing in the derivation of the Plefka expansion requires particular conditions on $\boldsymbol{K}$. A simple choice is then to suppose that $\boldsymbol{K}$ is a real matrix with elements that are randomly distributed Gaussian variables with zero mean and variance $\langle K_{ij}^2 \rangle = 1/N$, drawn independently except for the correlation

$$\langle K_{ji} K_{ij} \rangle = \frac{\eta}{N} \tag{2.4.1}$$

The parameter $\eta \in [-1, 1]$ controls the degree to which the matrix $\boldsymbol{K}$ is symmetric, i.e. it is a measure of symmetry for the physical couplings in the system. Such ensembles of matrices with Gaussian-distributed elements were first studied by Girko [41] and Ginibre [42]: their characteristic feature is that unless $\eta = 1$, the eigenvalues are not restricted to the real axis but distributed over an area in the complex plane.

For $\eta = 1$ we have symmetric matrices, which belong to what is known as the Wigner or Gaussian Orthogonal Ensemble. Symmetry here ensures that the dynamics obeys detailed balance with respect to the energy function $\sum_i \lambda x_i^2/2 + \sum_{ij} x_i K_{ij} x_j/2$ so that the stationary regime is an equilibrium

state.

The value $\eta = 0$ means that all correlations between matrix elements vanish and thus identifies a fully asymmetric $\boldsymbol{K}$: such random matrices, with completely independent real entries, belong to the Ginibre Orthogonal Ensembles [42]. Finally, $\eta = -1$ describes the antisymmetric case, where all eigenvalues of $\boldsymbol{K}$ lie along the imaginary axis because i$\boldsymbol{K}$ is Hermitian.

2.4.2 Extended Plefka Expansion

We next evaluate the predictions of the extended Plefka approach for our system with equation of motion (2.3.50). As shown in (2.3.55) above, the effective single-site dynamics is given by

$$\frac{dx_i(t)}{dt} = -\lambda x_i(t) + \sum_j K_{ij}\mu_j(t) + \int_0^t dt' \sum_j K_{ij}R_j(t,t')K_{ji}\delta x_i(t') + \phi_i(t) \tag{2.4.2}$$

where

$$\phi_i(t) = \xi_i(t) + \chi_i(t) \qquad \langle \phi_i(t)\phi_i(t')\rangle = \Sigma\delta(t-t') + \sum_j K_{ij}^2 \delta C_j(t,t') \tag{2.4.3}$$

The dynamics of the means $\mu_i(t)$, obtained by averaging over the ensemble (2.4.2) as in (2.3.56), is in full agreement with the exact one obtained by simply taking the mean of (2.3.50).

Let us calculate the response, which in the Plefka approach is given by (2.3.57). The dependence on the site i on the r.h.s. arises only from the term $\sum_j K_{ij}K_{ji}R_j(t,t')$. Because $K_{ij}K_{ji}$ is of order $1/N$, while the R_j are of order unity and are expected to have vanishing correlation with $K_{ij}K_{ji}$ (for any fixed i) for large N, this sum is self-averaging: for large N it can be replaced by

$$\sum_j K_{ij}K_{ji}R_j(t,t') \sim \frac{\eta}{N}\sum_j R_j(t,t') \equiv \eta R(t,t') \tag{2.4.4}$$

because $\langle K_{ij}K_{ji}\rangle = \eta/N$. For latter we note that the non-trivial term in the noise covariance (2.4.3) self-averages similarly to

$$\sum_j K_{ij}^2 C_j(t,t') \sim \frac{1}{N}\sum_j C_j(t,t') \equiv C(t,t') \tag{2.4.5}$$

The self-averaged version of (2.3.57) now reads

$$\frac{\partial R_i(t,t')}{\partial t} = -\lambda R_i(t,t') + \eta\int_{t'}^t dt'' R(t,t'')R_i(t'',t') + \delta(t-t') \tag{2.4.6}$$

From this one sees that all sites i will have the same response for large N, which makes sense because with our long-range disordered couplings all sites i become equivalent. We can thus drop the site index on R_i from now on, or formally average over i to get an equation for R.

As explained above we now consider the long-time limit where a steady state should be reached so that the response becomes TTI, $R(t,t') = R(t-t')$

$$\frac{\partial R(t-t')}{\partial t} = -\lambda R(t-t') + \eta \int_{t'}^{t} dt'' R(t-t'')R(t''-t') + \delta(t-t') \tag{2.4.7}$$

Laplace transforming with respect to time differences, with z the conjugate variable, gives

$$(z+\lambda)\tilde{R}(z) = \eta \tilde{R}^2(z) + 1 \tag{2.4.8}$$

This second order equation for the Laplace transformed response $\tilde{R}(z)$ has solution

$$\tilde{R}(z) = \frac{1}{2\eta}(z+\lambda) - \frac{1}{2\eta}\sqrt{(z+\lambda)^2 - 4\eta} \tag{2.4.9}$$

Here the sign is chosen to retrieve the correct behaviour for $z \to \infty$: as $R(t-t')$ must approach unity for small time differences, the Laplace transform $\tilde{R}(z)$ has to decay as $1/z$ for large z. The result (2.4.9) is particularly simple for $\eta = 0$, where the response takes the form

$$\tilde{R}(z) = \frac{1}{z+\lambda} \tag{2.4.10}$$

We next apply the same approach to the calculation of the connected correlations $\delta C(t,t')$. As we will only consider connected correlations in the following we drop the δ and write simply $C(t,t')$. We start from (2.3.58), make the self-averaging replacement (2.4.5), drop the site index and obtain

$$\frac{\partial C(t-t')}{\partial t} = -\lambda C(t-t') + \eta \int_{-\infty}^{t} dt'' R(t-t'')C(t''-t') + \int_{-\infty}^{t'} dt''[\Sigma\delta(t-t'') + C(t-t'')]R(t'-t'') \tag{2.4.11}$$

We take a two-sided Laplace transform of this

$$z\tilde{C}(z) = -\tilde{C}(z) + \eta\tilde{R}(z)\tilde{C}(z) + [\Sigma + \tilde{C}(z)]\tilde{R}(-z) \tag{2.4.12}$$

and solve to get

$$\tilde{C}(z) = \frac{\Sigma\tilde{R}(-z)}{z + \lambda - \tilde{R}(-z) - \eta\tilde{R}(z)} = \frac{\Sigma\tilde{R}(z)\tilde{R}(-z)}{1 - \tilde{R}(z)\tilde{R}(-z)} \tag{2.4.13}$$

In the second equality we have simplified using (2.4.8) to obtain a form that is manifestly even in z, as it should be because $C(t-t') = C(t'-t)$. Note that for the response, which is causal so vanishes for negative time differences, the two-sided Laplace transform reduces to the one-sided version.

2.4.3 Exact Solution

To assess whether the above predictions of the extended Plefka method are correct, we now study the exact solution of our model.

We will require as an essential ingredient the spectral density $\rho(k)$ of $\boldsymbol{K}$ in the thermodynamic limit, which follows from general theorems, namely Girko's elliptic and circular laws and the Wigner semicircular law. Girko's elliptic law [43] states that the average eigenvalue distribution $\rho(k)$ of $N \times N$ random matrices $\boldsymbol{K}$ drawn from a Gaussian ensemble described by (2.4.1), in the limit $N \to \infty$, is

$$\rho(k) = \begin{cases} \frac{1}{\pi(1-\eta^2)} & \left(\frac{x}{1+\eta}\right)^2 + \left(\frac{y}{1-\eta}\right)^2 < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.4.14}$$

where we have written x and y for the real and imaginary values of the eigenvalue k. The density $\rho(k)$ is uniform in an ellipse in the complex plane whose semi-axes are $1+\eta$ and $1-\eta$, respectively, along the real and imaginary directions, and whose foci are $\pm 2\sqrt{\eta}$. In the limit $\eta \to 1$ the Wigner semicircle law [36] is recovered from this, for the distribution of real eigenvalues of matrices from the Wigner ensemble

$$\rho(k) = \frac{1}{2\pi}\sqrt{4-k^2} \qquad k \in [-2, 2] \tag{2.4.15}$$

Girko's elliptic law can then be regarded as the generalization of Wigner's semicircular law to the case of an arbitrary degree of symmetry. For $\eta = 0$ the ellipse degenerates into the unit circle. Let us consider the vectorial form of the dynamics (2.3.50) of our model, where we temporarily add an external field $\boldsymbol{l}$ on the r.h.s.

$$\frac{d\boldsymbol{x}(t)}{dt} = -\lambda \boldsymbol{x}(t) + \boldsymbol{K}\boldsymbol{x}(t) + \boldsymbol{\xi}(t) + \boldsymbol{l}(t) \tag{2.4.16}$$

The solution can be written symbolically as, if we ignore contributions from the initial conditions

$$\boldsymbol{x}(t) = \int_0^t dt' \mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t')}[\boldsymbol{\xi}(t') + \boldsymbol{l}(t')] \tag{2.4.17}$$

This gives directly for the response function matrix

$$\boldsymbol{R}(t,t') = \left.\frac{\partial\langle \boldsymbol{x}(t)\rangle}{\partial \boldsymbol{l}(t')}\right|_{l=0} = \theta(t-t')\mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t')} \tag{2.4.18}$$

and we can set the field to zero again from now on. The responses are functions with step discontinuities and exponentially decaying with the time difference $t-t'$. The time sequence imposed by the θ-function implements the causality constraint, i.e. the response of the system at time t can be caused only by a perturbation at a previous time t'. One sees also that λ must be greater than the real part of all eigenvalues k of $\boldsymbol{K}$, to avoid exponentially increasing solutions. As the expression for $\boldsymbol{R}$ is TTI, it has a simple representation in the Laplace domain

$$\tilde{\boldsymbol{R}}(z) = \int_0^{+\infty} \mathrm{e}^{(\boldsymbol{K}-\lambda)s}\mathrm{e}^{-zs} ds = [z - (\boldsymbol{K}-\lambda)]^{-1} \qquad s \equiv t - t' \tag{2.4.19}$$

For comparison with the Plefka approach we are interested in $R(t-t') = (1/N)\sum_i R_{ii}(t-t') = \mathrm{Tr}\,\boldsymbol{R}(t-t')$ if we denote by Tr the normalized trace. This can be evaluated by integrating over the spectral density

$$\tilde{R}(z) = \langle \mathrm{Tr}\,\tilde{\boldsymbol{R}}(z)\rangle = \int dk\,\rho(k)[z-(k-\lambda)]^{-1} = \begin{cases} \frac{1}{2\eta}(\lambda+z) - \frac{1}{2\eta}\sqrt{(\lambda+z)^2-4\eta} & \eta \text{ generic} \\ \frac{1}{2}(\lambda+z) - \frac{1}{2}\sqrt{(\lambda+z)^2-4} & \eta = 1 \\ \frac{1}{z+\lambda} & \eta = 0 \end{cases} \tag{2.4.20}$$

Note that the expressions (2.4.20) are valid only for $z+\lambda$ outside the support of the eigenvalue spectrum as otherwise the integrand has singularities. Meaningful values can still be assigned to the integral for z inside the support, by appropriate regularization, and this is necessary when $\tilde{R}(z)$ is regarded as a resolvent from which spectral information is to be obtained, see e.g. Mehlig and Chalker [44] or Sommers et al. [45] for an interesting analogy with a two-dimensional classical electrostatic field calculation. We briefly sketch it. Let us start with the complex conjugate of the response, $\bar{\tilde{R}}(z)$, that can be written as

$$\bar{\tilde{R}}(z) = \left\langle \mathrm{Tr}(\bar{z}+\lambda-\boldsymbol{K})^{-1}\right\rangle = \int dk\rho(k)\frac{1}{\bar{z}+\lambda-k} = \int dk\rho(k)\frac{z+\lambda-k}{|z+\lambda-k|^2} = \frac{E(z+\lambda)}{2} \tag{2.4.21}$$

where $E(z+\lambda)$ is the 2d electrostatic field at position $z+\lambda$ generated by a charge distribution $\rho(k)$, as it is described in the complex plane. Therefore, if $\rho(k)$ is uniform on the unit circle, one can resort to well-known results for the electric field inside and outside a uniformly charged disk

$$\tilde{R}(z) = \frac{\bar{E}(z+\lambda)}{2} = \begin{cases} \frac{1}{z+\lambda} & |z+\lambda| \geq 1 \\ \bar{z}+\lambda & |z+\lambda| < 1 \end{cases} \tag{2.4.22}$$

The same analogy can be exploited also in the case of generic η and led Sommers et al. [45] to

$$\tilde{R}(z) = \begin{cases} \frac{1}{2\eta}(z+\lambda) - \frac{1}{2\eta}\sqrt{(z+\lambda)^2-4\eta} & |(z+\lambda) - \sqrt{(z+\lambda)^2-4\eta}| > 2\eta \\ \frac{(z_x+\lambda)}{(1+\eta)} - \mathrm{i}\frac{z_y}{1-\eta} & |(z+\lambda) - \sqrt{(z+\lambda)^2-4\eta}| < 2\eta \end{cases} \tag{2.4.23}$$

In our case $\tilde{R}(z)$ is a Laplace transform, as it was in the Plefka calculation, so we are only interested in its behaviour for large enough real z and the analytic continuation from this region, which is exactly what (2.4.20) provides. To be precise, (2.4.20) with the square root assigned its principal value is valid for $\mathrm{Re}(z) > -\lambda$, i.e. to the right of the midpoint of the branch cut between $z = -\lambda - 2\sqrt{\eta}$ and $z = -\lambda + 2\sqrt{\eta}$; to the left, one has to use the negative of the principal value to ensure that $\tilde{R}(z)$ is analytic except in the branch cut. Comparing with (2.4.9), we thus conclude that the extended Plefka method gives the exact response function for our system.

We next turn to the correlation function. To obtain the exact expressions for this we have to resort to different tools. Information about the spectrum is no longer enough, we also require the statistics of correlations between the left and right eigenvectors of $\boldsymbol{K}$; these eigenvectors are different in the generic case where $\boldsymbol{K}$ is not Hermitian, i.e. for $\eta \neq \pm 1$. Eigenvector statistics in non-Hermitian random matrix ensembles were studied extensively by Chalker and Mehlig [44] and we exploit their approach, slightly adjusted for our case of matrices with real rather than complex elements.

As for the response we start from the full non-local correlation matrix, which from (2.4.17) is given by

$$\begin{aligned}\boldsymbol{C}(t,t') = \langle \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t')\rangle &= \int_0^t \int_0^{t'} dt''dt''' \mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t'')}\langle \boldsymbol{\xi}(t'')\boldsymbol{\xi}^{\mathrm{T}}(t''')\rangle \mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})(t'-t''')} \\ &= \Sigma \int_0^{\min(t,t')} dt'' \mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t'')}\mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})(t'-t'')} \end{aligned} \tag{2.4.24}$$

In terms of the equal-time correlator

$$\boldsymbol{C}(t,t) = \Sigma \int_0^t d\tau\, \mathrm{e}^{(-\lambda+\boldsymbol{K})\tau}\mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})\tau} \tag{2.4.25}$$

this simplifies to

$$\boldsymbol{C}(t,t') = \begin{cases} \mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t')}\boldsymbol{C}(t',t') & t \geq t' \\ \boldsymbol{C}(t,t)\mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})(t'-t)} & t' > t \end{cases} \tag{2.4.26}$$

In the long-time limit $\boldsymbol{C}(t,t')$ will become TTI again, with $\boldsymbol{C}(t,t') = \boldsymbol{C}(t-t')$; $\boldsymbol{C}(0)$ then is the long-time limit of $\boldsymbol{C}(t,t)$. Combining the expressions for the two relative orderings of t and t' above and performing a two-sided Laplace transform with respect to the time difference gives

$$\begin{aligned}\tilde{\boldsymbol{C}}(z) &= \int_0^\infty ds\, \mathrm{e}^{-zs}\boldsymbol{C}(0)\mathrm{e}^{(-\lambda+\boldsymbol{K}^T)s} + \int_{-\infty}^0 ds\, \mathrm{e}^{-zs}\mathrm{e}^{(-\lambda+\boldsymbol{K})s}\boldsymbol{C}(0) = \\ &= \boldsymbol{C}(0)(-z+\lambda-\boldsymbol{K}^T)^{-1} + (z+\lambda-\boldsymbol{K})^{-1}\boldsymbol{C}(0)\end{aligned} \tag{2.4.27}$$

For further analysis it is useful to rewrite $\boldsymbol{C}(0)$ as

$$\begin{aligned}\boldsymbol{C}(0) \quad &= \Sigma \int_0^\infty d\tau_1 \int_0^\infty d\tau_2\, \mathrm{e}^{(-\lambda+\boldsymbol{K})\tau_1}\mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})\tau_2}\delta(\tau_1-\tau_2) = \\ &= \Sigma \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \int_{-\infty}^{+\infty} \tfrac{d\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}\omega(\tau_1-\tau_2)}\, \mathrm{e}^{(-\lambda+\boldsymbol{K})\tau_1}\, \mathrm{e}^{(-\lambda+\boldsymbol{K}^{\mathrm{T}})\tau_2}\end{aligned} \tag{2.4.28}$$

so that after integrating over τ_1 and τ_2 one has

$$\boldsymbol{C}(0) = \Sigma \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi}(\lambda - \mathrm{i}\omega - \boldsymbol{K})^{-1}(\lambda + \mathrm{i}\omega - \boldsymbol{K}^{\mathrm{T}})^{-1} \tag{2.4.29}$$

For a comparison with the $\tilde{C}(z)$ obtained in the Plefka approximation we need the normalized trace again, as in the case of the response, and combining (2.4.27) and (2.4.29) this takes the form

$$\mathrm{Tr}\,\tilde{\boldsymbol{C}}(z) = \Sigma \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi}\, \frac{1}{z-\mathrm{i}\omega}\Big\{\langle \mathrm{Tr}[(\lambda-z-\boldsymbol{K})^{-1}(\lambda+\mathrm{i}\omega-\boldsymbol{K}^{\mathrm{T}})^{-1}]\rangle - \langle \mathrm{Tr}[(\lambda-\mathrm{i}\omega-\boldsymbol{K})^{-1}(\lambda+z-\boldsymbol{K}^{\mathrm{T}})^{-1}]\rangle\Big\} \tag{2.4.30}$$

where the simple matrix identity

$$(a - \boldsymbol{A})^{-1} - (b - \boldsymbol{A})^{-1} = (a - \boldsymbol{A})^{-1}(b - \boldsymbol{A})^{-1}(b - a) \tag{2.4.31}$$

has been applied. We have explicitly added an average over the random sampling of $\boldsymbol{K}$ in order to be able to use random matrix technique for further evaluation. This is justified because like the response, which depends only on the spectrum and is self-averaging for large N because the spectrum is, the correlation function is also expected to be self-averaging. For later notational convenience we have also transformed $z \to -z$ on the r.h.s. of (2.4.30), anticipating that the final result (2.4.37) will be even in z.

The benefit of the above manipulations is that the calculation of the exact correlations is now reduced to finding the quadratic resolvents

$$\langle \mathrm{Tr}[(\lambda - \mathrm{i}\omega - \boldsymbol{K})^{-1}(\lambda + z - \boldsymbol{K}^{\mathrm{T}})^{-1}] \rangle \tag{2.4.32}$$

$$\langle \mathrm{Tr}[(\lambda - z - \boldsymbol{K})^{-1}(\lambda + \mathrm{i}\omega - \boldsymbol{K}^{\mathrm{T}})^{-1}] \rangle \tag{2.4.33}$$

Adapting the technique of [44] to our case of real-valued matrices, we find for such resolvents the general result

$$\langle \mathrm{Tr}[(z_1 - \boldsymbol{K})^{-1}(\bar{z}_2 - \boldsymbol{K}^{\mathrm{T}})^{-1}] \rangle = \frac{g_1 \bar{g}_2}{1 - g_1 \bar{g}_2} \tag{2.4.34}$$

where

$$g_1 = \frac{z_1 - \sqrt{z_1^2 - 4\eta}}{2\eta} \qquad \bar{g}_2 = \frac{\bar{z}_2 - \sqrt{\bar{z}_2^2 - 4\eta}}{2\eta} \tag{2.4.35}$$

Comparing with (2.4.20), one observes that g_1 and $\bar{g}_2$ are themselves response functions, with z_1 and $\bar{z}_2$ respectively replacing $z + \lambda$. In the case $\eta = 0$, the r.h.s. of (2.4.34) simplifies further to $1/(z_1 \bar{z}_2 - 1)$.

One expects the result (2.4.34) to apply whenever both z_1 and $\bar{z}_2$ are outside of the spectral ellipse. This is easily verified: one checks that $|g_1| = 1$ is another parametrization for the boundary of this ellipse

$$|g_1| = 1 \Leftrightarrow \quad \left| z_1 - \sqrt{z_1^2 - 4\eta} \right| = 2\eta \tag{2.4.36}$$

with foci $z_1 = \pm 2\sqrt{\eta}$ and semi-axes $1+\eta$ and $1-\eta$ as before. So z_1 and $\bar{z}_2$ are outside of the spectral ellipse when $|g_1| < 1$ and $|\bar{g}_2| < 1$, which ensures that (2.4.34) is non-singular. The resolvent then diverges when e.g. $z_1 = z_2$ and z_1 approaches the boundary of the ellipse.

To work out the trace (2.4.30) defining the Laplace transformed correlation function, we need to set in the first resolvent (2.4.32) $z_1 = \lambda - \mathrm{i}\omega$ and $\bar{z}_2 = \lambda + z$, and in the second (2.4.33) $z_1 = \lambda - z$

and $\bar{z}_2 = \lambda + \mathrm{i}\omega$. After these substitutions the integration can be conveniently carried out using residues (see appendix A), with the result

$$\tilde{C}(z) = \frac{\Sigma\, \tilde{R}(z)\tilde{R}(-z)}{1 - \tilde{R}(z)\tilde{R}(-z)} \tag{2.4.37}$$

This is identical to the prediction (2.4.13) of the Plefka approximation. Our conclusion is, therefore, that for our model with weak long-range interactions the extended Plefka approach provides fully exact results for response and correlation functions, in the thermodynamic limit $N \to \infty$.

2.5 Quantitative results

In this section we look at the quantitative results for our model system in more detail. As the model does not obey detailed balance when $\eta < 1$, we are in general dealing with a non-equilibrium steady state and will see some non-trivial features emerge from this.

We focus initially on the correlation function (2.4.37), which after substituting $\tilde{R}(z)$ and $\tilde{R}(-z)$ and simplifying reads

$$\tilde{C}(z) = \frac{4\,\Sigma}{[(\lambda + z) + \sqrt{(\lambda + z)^2 - 4\eta}][(\lambda - z) + \sqrt{(\lambda - z)^2 - 4\eta}] - 4} \tag{2.5.1}$$

Particular cases of note are

$$\tilde{C}(z)\big|_{\eta=0} = \frac{\Sigma}{\lambda^2 - z^2 - 1} \tag{2.5.2}$$

$$\tilde{C}(z)\big|_{\eta=1} = \Sigma\Big[-\frac{1}{2} + \frac{1}{4z}\sqrt{(\lambda + z)^2 - 4} - \frac{1}{4z}\sqrt{(\lambda - z)^2 - 4}\Big] \tag{2.5.3}$$

$$\tilde{C}(z)\big|_{\eta=-1} = \Sigma\Big[-\frac{1}{2} + \frac{1}{4\lambda}\sqrt{(\lambda + z)^2 + 4} + \frac{1}{4\lambda}\sqrt{(\lambda - z)^2 + 4}\Big] \tag{2.5.4}$$

where the middle one is the detailed balance limit.

The long-time behaviour of $C(t - t')$ is determined by the singularities, i.e. poles and branch cuts, of $\tilde{C}(z)$ that are closest to the origin. It will be useful to think of these in relation to two copies of the spectral ellipse: bearing in mind that $\tilde{R}(\pm z) = g_1(z \mp \lambda)$, these are shifted to have their centres at $\pm\lambda$.

For generic η (see figures A.2, A.4 in appendix A), each of the two square roots in $\tilde{C}(z)$ contributes a branch cut[1]. Each branch cut lies completely inside the relevant shifted spectral ellipse, and

[1] In this context branch cuts correspond to the lines in the complex plane for which the Laplace transform is not well-defined. In fact, all the values of z such that the argument of a square root becomes negative constitute a line of discontinuity: either +i or −i can be extracted from the square root and this jump makes the function multivalued.

extends from one focus of the ellipse to the other. Explicitly, the branch cuts are

$$\pm\lambda - 2\sqrt{\eta} < \mathrm{Re}(z_{\mathrm{bc}}) < \pm\lambda + 2\sqrt{\eta} \qquad \mathrm{Im}(z_{\mathrm{bc}}) = 0 \qquad \eta > 0 \tag{2.5.5a}$$

$$\mathrm{Re}(z_{\mathrm{bc}}) = \pm\lambda \qquad -2\sqrt{|\eta|} < \mathrm{Im}(z_{\mathrm{bc}}) < +2\sqrt{|\eta|} \qquad \eta < 0 \tag{2.5.5b}$$

In the symmetric and anti-symmetric limits the ellipses degenerate to straight lines that coincide with the branch cuts while in the asymmetric case $\eta = 0$ each branch cut shrinks to a point $z_{\mathrm{bc}} = \pm\lambda$ at the centre of the spectral circle; see figure A.1 in appendix A.

In addition to branch cuts, the Laplace transformed correlation function (2.5.1) can have poles (for $\eta \neq 1, -1$). Setting the denominator of (2.5.1) to zero gives

$$z_{\mathrm{pole}} = \pm z_0 \;\text{ with }\; z_0 = \left(\frac{1-\eta}{1+\eta}\right)\sqrt{\lambda^2 - (1+\eta)^2} \tag{2.5.6}$$

These poles emerge from the branch cuts as λ is decreased below the threshold value $\lambda_{\mathrm{threshold}} = (1+\eta)^2/(2\sqrt{\eta})$ for $\eta > 0$ and $\lambda_{\mathrm{threshold}} = (1-\eta^2)/(2\sqrt{|\eta|})$ for $\eta < 0$ (see figure 2.1); they do not exist for larger λ because they are then no longer on the physical branch of $\tilde{C}(z)$. With decreasing λ they then move towards the origin and reach it at a critical value for λ given by $\lambda_{\mathrm{min}}(\eta) = 1 + \eta$. This makes sense as the largest real part of eigenvalues within the spectral ellipse of $\boldsymbol{K}$ is exactly $1 + \eta$: for $\lambda < \lambda_{\mathrm{min}}$ these eigenvalues would cause the correlation function to diverge for long time differences.

2.5.1 Existence of poles

The denominator in (2.5.1) is zero when

$$\frac{1}{2}\left[(\lambda + z) + \sqrt{(\lambda+z)^2 - 4\eta}\right] \cdot \frac{1}{2}\left[(\lambda - z) + \sqrt{(\lambda - z)^2 - 4\eta}\right] = 1 \tag{2.5.7}$$

That corresponds to

$$\tilde{R}(z)\tilde{R}(-z) = 1 \tag{2.5.8}$$

As we want to define the condition of existence for z_{pole} that lies on the real axis, we shall consider z real.

1. $z > 0$: $\tilde{R}(z) > 1$. For (2.5.7) to be true, we need to consider the values of z for which $\tilde{R}(-z) < 1$. We know (as can be verified from the explicit expression) that $|\tilde{R}(-z)|$ (which is equal to $\tilde{R}(-z)$ for real z) is smaller than 1 outside the ellipse intersecting the real axis in $z = \lambda - (1+\eta)$. Thus, for z positive, the pole exists only if $z > \lambda - (1+\eta)$

2. $z < 0$: $\tilde{R}(-z) > 1$. By a similar argument, the condition of existence is given by $z < -\lambda + (1+\eta)$.

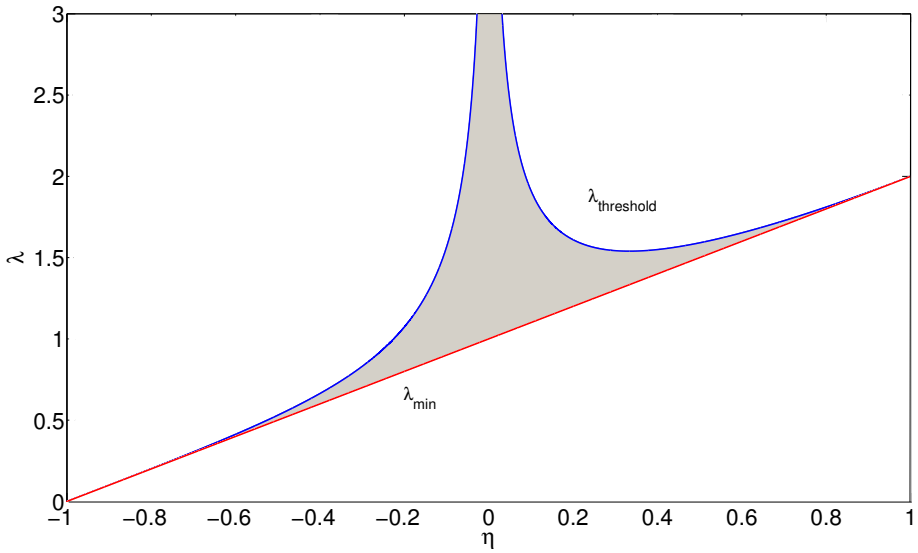

Figure 2.1: $\lambda_{\text{threshold}}$ and λ_{min} as a function of η. The pole exists for $\lambda_{\text{min}} < \lambda < \lambda_{\text{threshold}}$, i.e. in the shaded area.

When z_{pole} is defined, one has $z_{\text{pole}} < z_{\text{bc}}$. Let us set $z_{\text{min}} = \lambda - (1+\eta)$. Thus one has $z_{\text{min}} < z_{\text{pole}} < z_{\text{bc}}$. An analogous condition can be written for the pole at negative z. One can look at the values of λ, as a function of η, for which the pole emerges. The minimum value λ_{min} is defined by setting $z_{\text{min}} = z_{\text{pole}}$ and it corresponds to $\lambda_{\text{min}} = 1+\eta$. The maximum value, $\lambda_{\text{threshold}}$ is found by $z_{\text{pole}} = z_{\text{bc}}$ and one has $\lambda_{\text{threshold}} = (1+\eta)^2/(2\sqrt{\eta})$ for $\eta > 0$. Thus z_{pole} is defined for $\lambda_{\text{min}} < \lambda < \lambda_{\text{threshold}}$. The analysis on the condition of existence is valid irrespective of the sign of η. We stress anyway that for $\eta < 0$ the branch cuts shrink to two points $z_{\text{bc}} = \pm\lambda$ and $z_{\text{pole}} < z_{\text{bc}}$. As a consequence $\lambda_{\text{threshold}}$ is given by setting $z_{\text{pole}} = \lambda$ and $\lambda_{\text{threshold}} = (1-\eta^2)/(2\sqrt{|\eta|})$ for $\eta < 0$.

2.5.2 Terminal decay rate of correlation function

The long-time or terminal decay rate r of the correlation function is now given by the singularity, be it pole or branch cut edge, that has the smallest (positive) real part, i.e. $r = \min(r_{\text{pole}}, r_{\text{bc}})$, and its inverse $1/r$ is the largest relaxation time. The real part of the pole is z_0 itself, $r_{\text{pole}} = z_0$, while for the branch cut it is, from (2.5.5a), $r_{\text{bc}} = \lambda - 2\sqrt{\eta}$ for $\eta > 0$ and $r_{\text{bc}} = \lambda$ otherwise. For $\lambda_{\text{min}} < \lambda < \lambda_{\text{threshold}}$, i.e. for which the pole exists, $r_{\text{pole}} < r_{\text{bc}}$ thus r_{pole} sets r, while for all other values of λ r_{bc} becomes responsible for the asymptotic decay. Remarkably, at $\eta = 0$, one has $r_{\text{pole}} = \sqrt{\lambda^2 - 1} < r_{\text{bc}} = \lambda$ so here the pole is relevant. Note that as $r_{\text{bc}} \leq \lambda$ for all η, we also have $r \leq \lambda$. Bearing in mind that for a non-interacting system we would have $C(t-t') = \Sigma \exp(-\lambda|t-t'|)$, this means that the asymptotic decay rate is only made smaller by the interactions, never larger.

Plotting r_{pole} and r_{bc} against η, one can verify that $r = r_{\text{pole}}$ for the η values, at some given λ, that

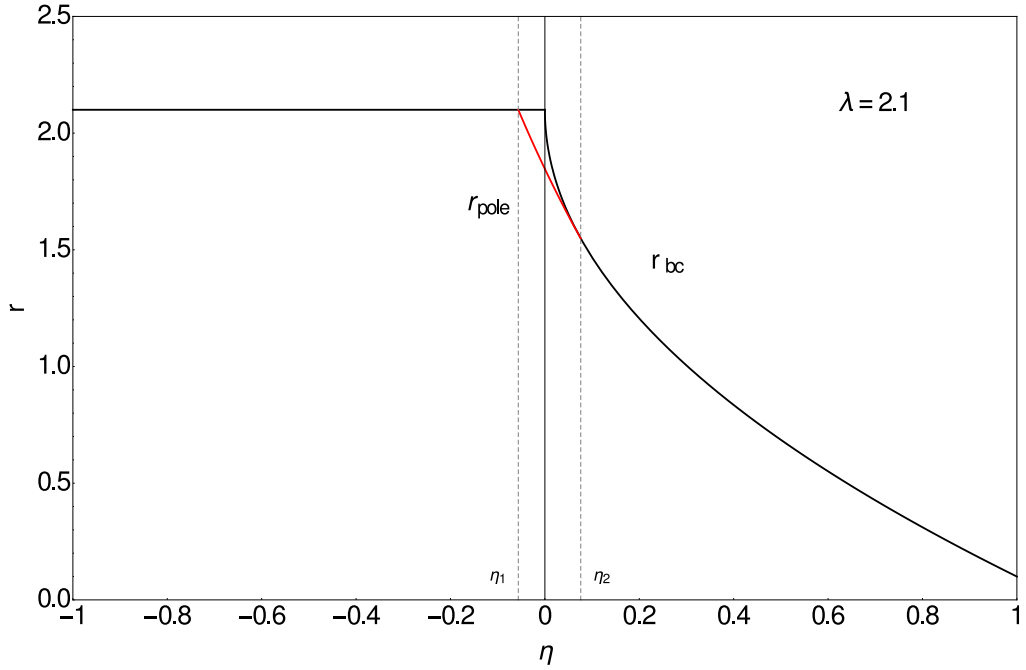

Figure 2.2: Typical plot for $\lambda > 2$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. It is defined by r_{pole} (red line) only for $\eta_1 < \eta < \eta_2$ as explained in case 1, otherwise it is set by r_{bc} (black line). $\lambda = 2.1$, $\eta_1 = -0.056$ and $\eta_2 = 0.076$.

satisfy the condition of existence for the poles - for all other η, $r = r_{\text{bc}}$. The η range for which the pole exists varies with λ because of the requirement that $\lambda_{\min}(\eta) < \lambda < \lambda_{\text{threshold}}(\eta)$ previously discussed. In particular, one can define 3 different cases, depending on λ, of how this change over in the singularity that is relevant for the asymptotics, from pole to branch cut and vice versa, occurs.

1. Let us consider $\lambda > 2$ (thus $\lambda > \lambda_{\min}$ for any η) as in figure 2.2. The pole exists for $\eta_1 < \eta < \eta_2$. η_1 and η_2 are given by setting $\lambda = \lambda_{\text{threshold}}(\eta)$ for positive and negative η: η_1 is defined in such a way that $(1 - \eta_1^2)/2\sqrt{|\eta_1|} = \lambda$ and η_2 is such that $\lambda = (1 + \eta_2)^2/2\sqrt{\eta_2}$.

2. The function $\lambda_{\text{threshold}}$ for $\eta > 0$ has a minimum $\lambda_{\text{threshold}} = 8/3\sqrt{3}$ at $\eta = 1/3$, thus we look at $8/3\sqrt{3} < \lambda < 2$ (see figure 2.3). The pole exists for η close to 0 and 1, more precisely for $\eta_{01} < \eta < \eta_{02}$ and $\eta_{11} < \eta < \eta_{12}$. We set η_{02} and η_{11} as the two solutions of $\lambda = (1 + \eta)^2/2\sqrt{\eta}$, η_{01} is defined by $(1 - \eta_{01}^2)/2\sqrt{|\eta_{01}|} = \lambda$ while η_{12} by $\lambda = \lambda_{\min} = 1 + \eta_{12}$.

3. Finally we take $\lambda < 8/3\sqrt{3}$, as graphically represented in figure 2.4. The pole exists for $\eta_1 < \eta < \eta_2$, where η_1 satisfies $(1 - \eta_1^2)/2\sqrt{|\eta_1|} = \lambda$ and η_2 is given by $\lambda = 1 + \eta_2$.

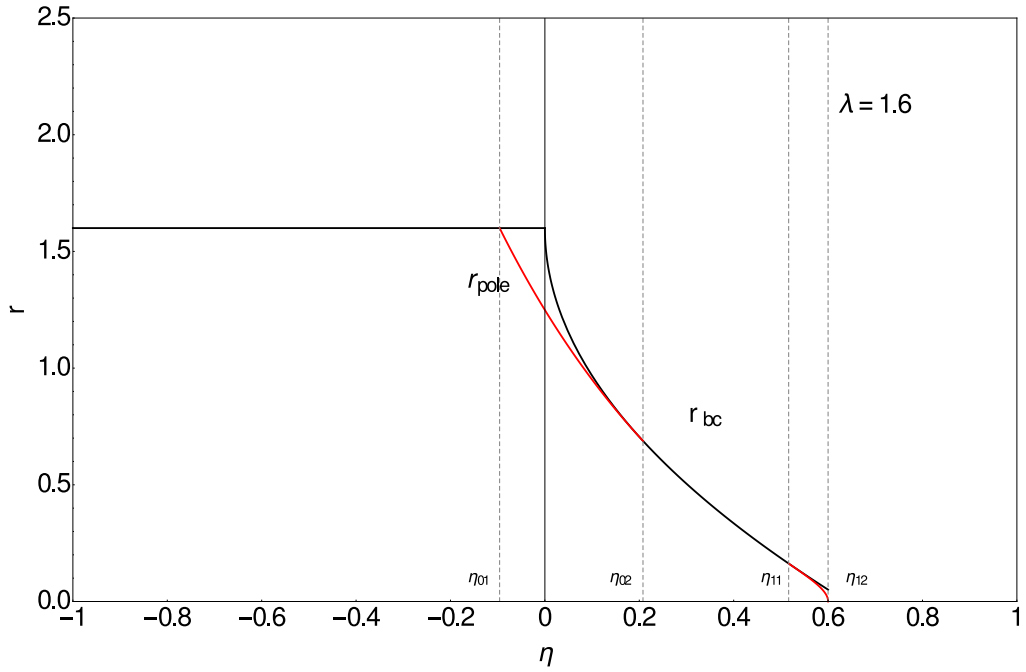

Figure 2.3: Typical plot for $8/3\sqrt{3} < \lambda < 2$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. It is defined by r_{pole} (red line) only for $\eta_{01} < \eta < \eta_{02}$ and $\eta_{11} < \eta < \eta_{12}$ as explained in case 2, otherwise it is set by r_{bc} (black line). $\lambda = 1.6$, $\eta_{01} = -0.096$, $\eta_{02} = 0.208$, $\eta_{11} = 0.517$ and $\eta_{12} = 0.6$. We plot r for the range of η where $\lambda_{\min} < \lambda$, i.e. $-1 \leq \eta < 0.6$, as the rest is unphysical.

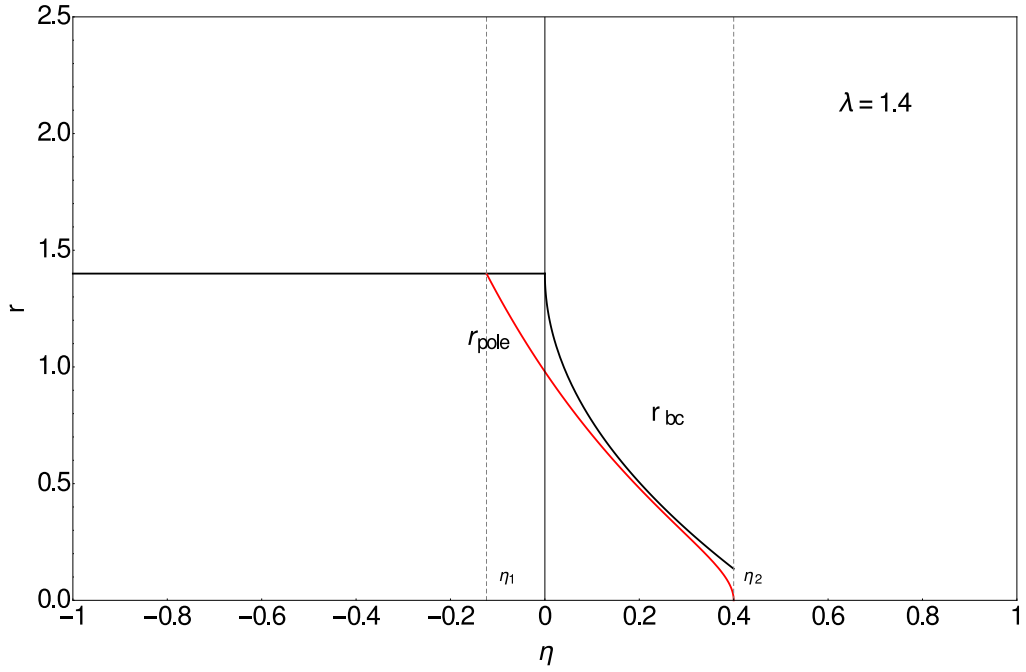

Figure 2.4: Typical plot for $\lambda < 8/3\sqrt{3}$ of the terminal decay $r = \min(r_{\text{bc}}, r_{\text{pole}})$ as a function of η. It is defined by r_{pole} (red line) only for $\eta_1 < \eta < \eta_2$ as explained in case 3, otherwise it is set by r_{bc} (black line). $\lambda = 1.4$, $\eta_1 = -0.124$ and $\eta_2 = 0.4$. We plot r for the values of η for which this λ is physical, i.e. $\lambda_{\min} < \lambda$ for $-1 \leq \eta < 0.4$.

2.5.3 Power Spectra and Power Laws

We can obtain the power spectrum of the fluctuations in our system by setting $z = \mathrm{i}\omega$ in (2.5.1), which converts the two-sided Laplace transform to a Fourier transform. For notational simplicity we use the same symbol $\tilde{C}(\omega)$ for the later as for the former, the meaning being clear from the argument of the function. Of primary interest is how the power spectrum differs from the simple Lorentzian case corresponding to a purely exponential correlation function decay.

We note first that the asymmetric case $\eta = 0$ in (2.5.2) always gives a Lorentzian power spectrum $\tilde{C}(\omega) = \Sigma/(\lambda^2 - 1 + \omega^2)$. The presence of the interactions only manifests itself here in a change of the characteristic frequency from λ to $r_{\mathrm{pole}} = \sqrt{\lambda^2 - 1}$. More generally for large λ any non-trivial features of the correlation function will be hidden underneath a rapidly decaying $\exp(-\lambda|t - t'|)$ envelope, giving a Lorentzian power spectrum. This can be seen formally by taking $\lambda \to \infty$ in (2.5.1) at z of order λ.

Non-trivial power spectra are then expected to appear in the opposite regime of small λ, or more precisely small $\lambda - \lambda_{\mathrm{min}}$ where $\lambda_{\mathrm{min}} = 1 + \eta$. Keeping the self-interaction in the vicinity of this critical value allows one to detect interesting features such as power law tails, as illustrated in figure 2.5. To make the comparison of different spectral shapes easier it is convenient to remove uninteresting prefactors, i.e. to extract the overall scales of $\tilde{C}(\omega)$ and ω and plot the normalized quantities. For $\tilde{C}(\omega)$ we take as the scale

$$C(0) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi}\, \tilde{C}(\omega) \tag{2.5.9}$$

A scale for ω can be extracted as the inverse of a typical timescale τ for the decay of correlations; we choose in particular a root mean squared decay time

$$\tau^2 = \frac{\int_{-\infty}^{+\infty} dt\, t^2 C(t)}{2\int_{-\infty}^{+\infty} dt\, C(t)} = \frac{\int_{-\infty}^{+\infty} dt\, t^2 C(t)}{2\tilde{C}(0)} \tag{2.5.10}$$

The factor 2 is needed in order to have a correct equality when the formula refers to correlations that are pure exponentials. It can be seen as an average over a distribution of relaxation times: this superposition of modes mixes different contributions and picks up rates faster than the termination decay $1/|z_{\mathrm{pole}}|$. Proceeding by some manipulations one has

$$\begin{aligned}
\int_{-\infty}^{+\infty} dt\, t^2 C(t) &= \int_{-\infty}^{+\infty} dt\, t^2 \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \tilde{C}(\omega) e^{-\mathrm{i}\omega t} = \\
&= -\int_{-\infty}^{+\infty} dt \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \tilde{C}(\omega) \frac{d^2}{d^2\omega} e^{-\mathrm{i}\omega t} = \\
&= -\int_{-\infty}^{+\infty} dt \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{d^2\tilde{C}(\omega)}{d^2\omega} e^{-\mathrm{i}\omega t} = \\
&= -\int_{-\infty}^{+\infty} d\omega \frac{d^2\tilde{C}(\omega)}{d^2\omega} \delta(\omega) = -\left.\frac{d^2\tilde{C}(\omega)}{d^2\omega}\right|_{\omega=0}
\end{aligned} \tag{2.5.11}$$

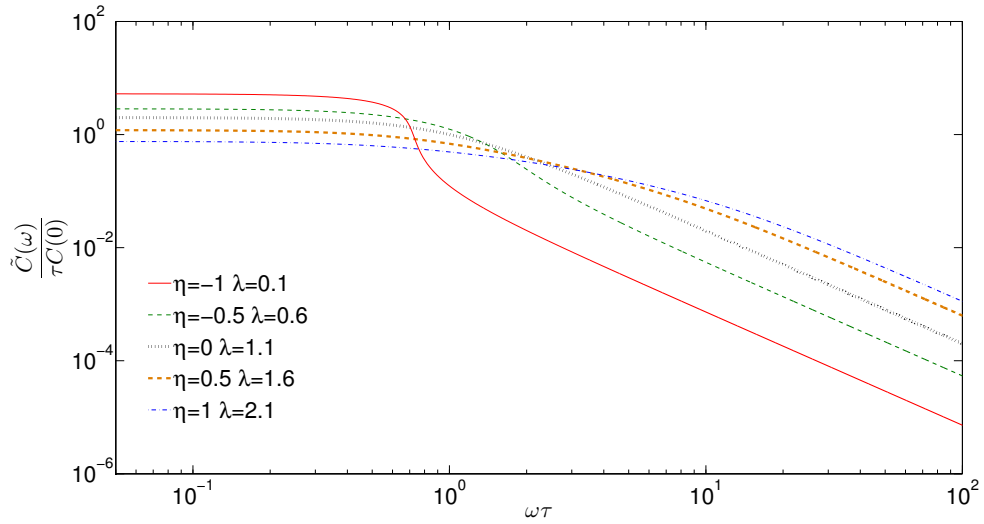

Figure 2.5: Log-log plots of normalized (see text) power spectra for different symmetries. λ is taken close to the corresponding minimal value $1 + \eta$ to highlight non-Lorentzian features. For small ω, the horizontal plateau represents an exponential cutoff, while the large ω tail $\sim 1/\omega^2$ is as for a Lorentzian ($\eta = 0$). The power spectra for $\eta > 0$ are broader than Lorentzian, suggesting slower decays that approach power laws for $\lambda \to 1 + \eta$. For $\eta < 0$ sharp drops in the power spectrum suggest oscillatory correlation decay in the time domain.

We then plot $\tilde{C}(\omega)/[\tau C(0)]$ versus $\omega\tau$ to ensure the normalized spectrum has a unit area under the curve. A log-log plot as in figure 2.5 shows clearly the large-frequency Lorentzian tail and suggests slower power law correlation decays for positive η and oscillatory decay for negative η.

We want to investigate more formally the emergence of power law behaviours for large time. This requires minimizing the effect of the exponential cut off provided by the self-interaction, so we consider $\lambda = \lambda_{\min}$. We then need to study the behaviour of $\tilde{C}(\omega)$ for small ω. For $\eta = 1$ one finds, by expansion of (2.5.3), $\tilde{C}(\omega) \sim 1/\sqrt{2\omega}$, corresponding to a $|t - t'|^{-1/2}$ decay in the time domain.

To understand the effect of slight deviations from symmetry we set $\eta = 1 - \epsilon$ with ϵ small. At fixed frequencies $\omega \sim O(1)$, the limit $\epsilon \to 0$ then just retrieves $\tilde{C}(\omega)|_{\eta=1}$, so the latter

$$\tilde{C}(\omega)|_{\eta=1} = -\frac{1}{2} + \frac{1}{2\sqrt{2}|\omega|}\sqrt{4 - \lambda^2 + \omega^2 + \sqrt{\omega^4 + 2\omega^2(\lambda^2+4) + (\lambda^2-4)^2}} \tag{2.5.12}$$

evaluated at $\lambda = \lambda_{\min} = 2$ is the limiting "master curve" for small ϵ in this part of the power spectrum

$$\tilde{C}(\omega)|_{\eta=1,\,\lambda=2} = -\frac{1}{2} + \frac{1}{2\sqrt{2}|\omega|}\sqrt{\omega^2 + \sqrt{\omega^4 + 16\omega^2}} \tag{2.5.13}$$

This master curve has asymptotic behaviour $\sim 1/\sqrt{2\omega}$ for small ω, as found above: the power spectrum for small ϵ generically contains a non-Lorentzian power law regime as our initial numer-

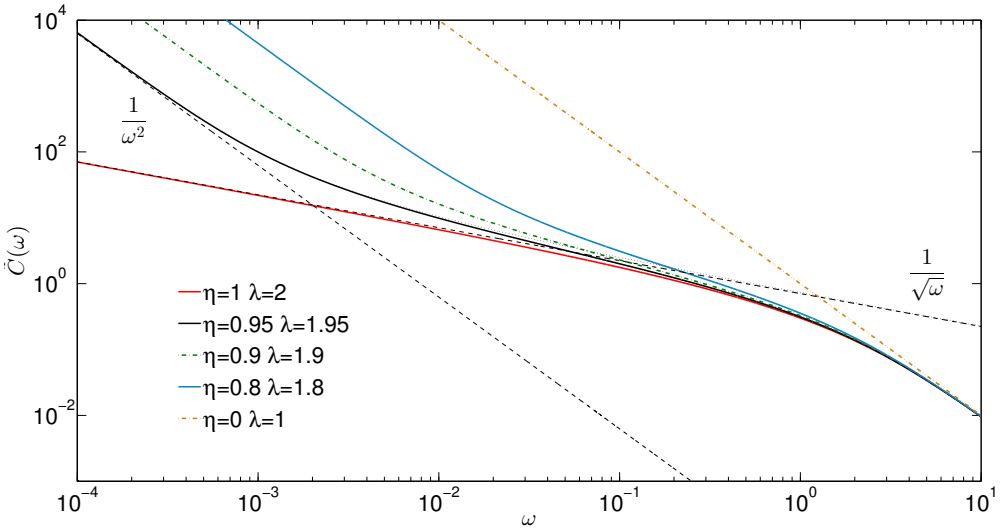

Figure 2.6: Power spectrum at minimal $\lambda = 1 + \eta$ for positive symmetry parameters η. Dashed lines show the asymptotic power laws at small frequency, which govern the long-time behaviour. For slight asymmetry ($\eta = 1 - \epsilon$), one sees interpolation between two master curves governing the frequency regimes of $\omega = O(1)$ and $\omega \sim \epsilon^2$. All curves show unnormalized power spectra, for noise amplitude $\Sigma = 1$.

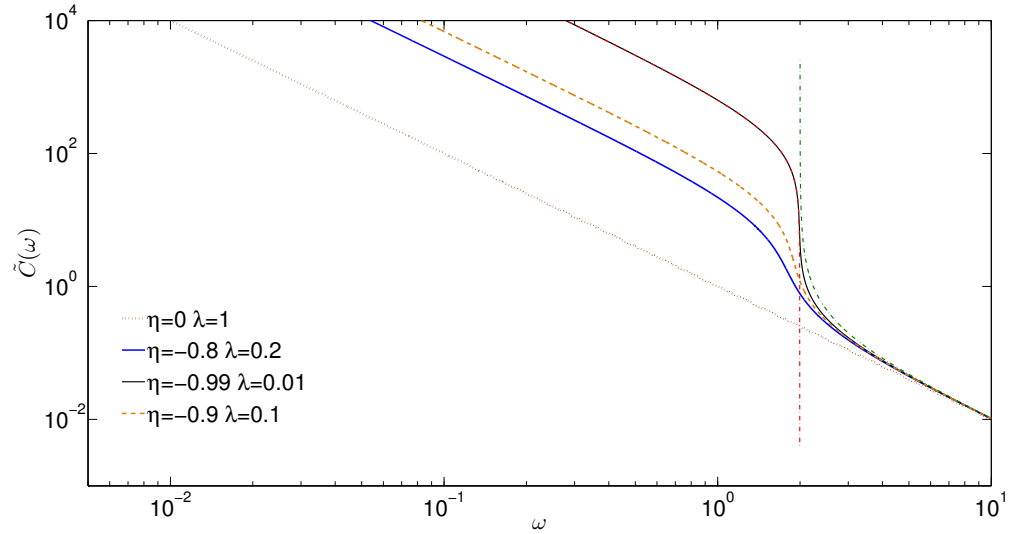

Figure 2.7: Analogue of figure 2.6 for negative symmetry parameters η. For small deviations from anti-symmetry ($\eta = -1 + \epsilon$) the power spectrum splits into two regimes at $\omega = 2$, each with its own master curve (dashed lines). The amplitude in the low frequency part diverges as $1/\epsilon$ while the higher frequencies have a finite amplitude for $\epsilon \to 0$, so that an effective frequency cutoff at $\omega = 2$ develops in the limit.

ics suggested.

If rather than fixing ω first and then taking $\epsilon \to 0$, we directly expand $\tilde{C}(\omega)$ for small ω at fixed ϵ, we find $\tilde{C}(\omega) \sim \frac{\epsilon^3}{2\omega^2}$ instead of $1/\sqrt{2\omega}$. Comparing the two expressions suggests that there is a crossover between two different regimes at a frequency scaling as ϵ^2. To analyze the crossover region we therefore set $\omega = \epsilon^2\gamma$ and take $\epsilon \to 0$ at fixed γ. The rescaled correlation $\epsilon\tilde{C}(\omega)$ then approaches a separate master curve

$$\hat{C}(\gamma) = \frac{1 + \sqrt{1+16\gamma^2} + 2\sqrt{2}\gamma\sqrt{-1+\sqrt{1+16\gamma^2}} + \sqrt{2}\sqrt{1+\sqrt{1+16\gamma^2}}}{8\gamma^2\epsilon} \tag{2.5.14}$$

The two tails of this low-frequency master curve retrieve the scalings found above as they should

$$\gamma \ll 1 \qquad (\omega \ll \epsilon^2) \qquad \hat{C}(\gamma) \sim 1/2\epsilon\gamma^2 \qquad \tilde{C}(\omega) \sim \epsilon^3/2\omega^2 \tag{2.5.15a}$$

$$\gamma \gg 1 \qquad (\omega \gg \epsilon^2) \qquad \hat{C}(\gamma) \sim 1/\epsilon\sqrt{2\gamma} \qquad \tilde{C}(\omega) \sim 1/\sqrt{2\omega} \tag{2.5.15b}$$

The results of the above analysis are illustrated in figure 2.6. Dashed lines indicate the exponents of the limiting power laws.

One notable aspect of the above power spectra is the $1/\omega^2$ tail for $\omega \to 0$, which makes the time-domain correlation function $C(t-t')$, obtained by inverse Fourier transform, formally infinite. This divergence could be regularized by taking λ slightly larger than $\lambda_{\min}$; it turns out that in this limit the dominant contribution to $C(t-t')$ is from the pole z_{pole} defined in (2.5.6). This contribution is of the order of $z_{\text{pole}}^{-1}\exp[-z_{\text{pole}}(t-t')]$, with z_{pole} scaling as $(\lambda - \lambda_{\min})^{1/2}$ (see appendix B for the detailed calculation).

Finally we consider the opposite end of the η range and study the case of a slight deviation from antisymmetry, given by $\eta = -1 + \epsilon$. To obtain the asymptotic behaviour for small ϵ, we expand the power spectrum in ϵ and retain the two leading orders (which are $O(\epsilon^{-1})$ and $O(\epsilon^0)$). This yields

$$\tilde{C}(\omega) \sim \frac{1}{\epsilon\omega^2}\left(4 + 2\sqrt{4-\omega^2} - \omega^2\right) + O(1) \qquad \omega < 2 \tag{2.5.16}$$

$$\tilde{C}(\omega) \sim \frac{1}{\omega^2 - 4} + O(\epsilon) \qquad \omega > 2 \tag{2.5.17}$$

and these limiting curves are shown as dashed lines in figure 2.7. The key observation is that for small ϵ the power spectrum is confined almost entirely to the frequency range $0 < \omega < 2$, while higher frequencies are suppressed relative to this by a factor of ϵ. As $\epsilon \to 0$, a hard frequency cutoff therefore emerges at $\omega = 2$.

2.5.4 Time Domain

To gain further insight we can extract analytically the exact correlations in the time domain for $\eta = 1$ (symmetric couplings) and $\eta = -1$ (anti-symmetric couplings). For the symmetric case, using $\boldsymbol{K}^{\mathrm{T}} = \boldsymbol{K}$ in (2.4.24) gives

$$C(t,t') = \Sigma \int_0^{\min(t,t')} dt''\, \mathrm{Tr}\, \mathrm{e}^{(-\lambda+\boldsymbol{K})(t+t'-2t'')} \tag{2.5.18}$$

The trace can be written as an integral over eigenvalues distributed according to Wigner's semi-circular law to give

$$\begin{aligned} C(t,t') &= \Sigma \int_{-2}^{2} \frac{dk}{2\pi} \sqrt{4-k^2} \int_0^{\min(t,t')} dt''\, \mathrm{e}^{(-\lambda+k)(t+t'-2t'')} = \\ &= \Sigma \int_0^{\min(t,t')} dt''\, \frac{I_1(2(t+t'-2t''))}{t+t'-2t''} \mathrm{e}^{-\lambda(t+t'-2t'')} = \\ &= \Sigma \int_{|t-t'|}^{t+t'} dw\, \frac{I_1(2w)}{2w} \mathrm{e}^{-\lambda w} \end{aligned} \tag{2.5.19}$$

In the first step, we changed variable $k = 2\cos\theta$ to write the k-integral as a modified Bessel function $\frac{I_1(\tau)}{\tau} = \frac{1}{\pi}\int_0^{\pi} d\theta (\sin\theta)^2 \mathrm{e}^{\tau\cos\theta}$. The final equality follows by setting $w = t + t' - 2t''$. In the long time limit the integral runs up to $+\infty$ and the result is manifestly TTI. The Fluctuation-Dissipation Theorem (FDT) [46] is then expected to hold because for symmetric couplings the system has detailed balance. This can be checked by calculating the response function, which comes out as simply the integrand of (2.5.19)

$$R(t-t') = \theta(t-t') \frac{I_1(2(t-t'))}{t-t'} \mathrm{e}^{-\lambda(t-t')} \tag{2.5.20}$$

From (2.4.18) one has

$$\begin{aligned} R(t-t') &= \mathrm{Tr}\, \boldsymbol{R}(t-t') = \theta(t-t') \mathrm{Tr}\, \mathrm{e}^{(-\lambda+\boldsymbol{K})(t-t')} = \\ &= \theta(t-t') \int_{-2}^{2} \frac{dk}{2\pi} \sqrt{4-k^2}\, \mathrm{e}^{(-\lambda+k)(t-t')} = \theta(t-t') \frac{I_1(2(t-t'))}{t-t'} \mathrm{e}^{-\lambda(t-t')} \end{aligned} \tag{2.5.21}$$

This is as expected from the FDT $TR(t-t') = -(\partial/\partial t) C(t-t')$ where in our case $T = \Sigma/2$. In fact in the long time limit, for e.g. $t > t'$

$$\begin{aligned} -\frac{\partial C(t-t')}{\partial t} &= \theta(t-t') \frac{\partial}{\partial t} \Big(\Sigma \int_{+\infty}^{t-t'} dw\, \frac{I_1(2w)}{2w} \mathrm{e}^{-\lambda w} \Big) = \\ &= \frac{\Sigma}{2} \theta(t-t') \frac{I_1(2(t-t'))}{(t-t')} \mathrm{e}^{-\lambda(t-t')} = \frac{\Sigma}{2} R(t-t') \end{aligned} \tag{2.5.22}$$

The power law behaviour we found above in Fourier space corresponds to a power law in the time domain as can be confirmed using the asymptotic expression of the modified Bessel function

$$I_1(z) \sim \frac{\mathrm{e}^z}{\sqrt{2\pi z}} \qquad z \gg 1 \tag{2.5.23}$$

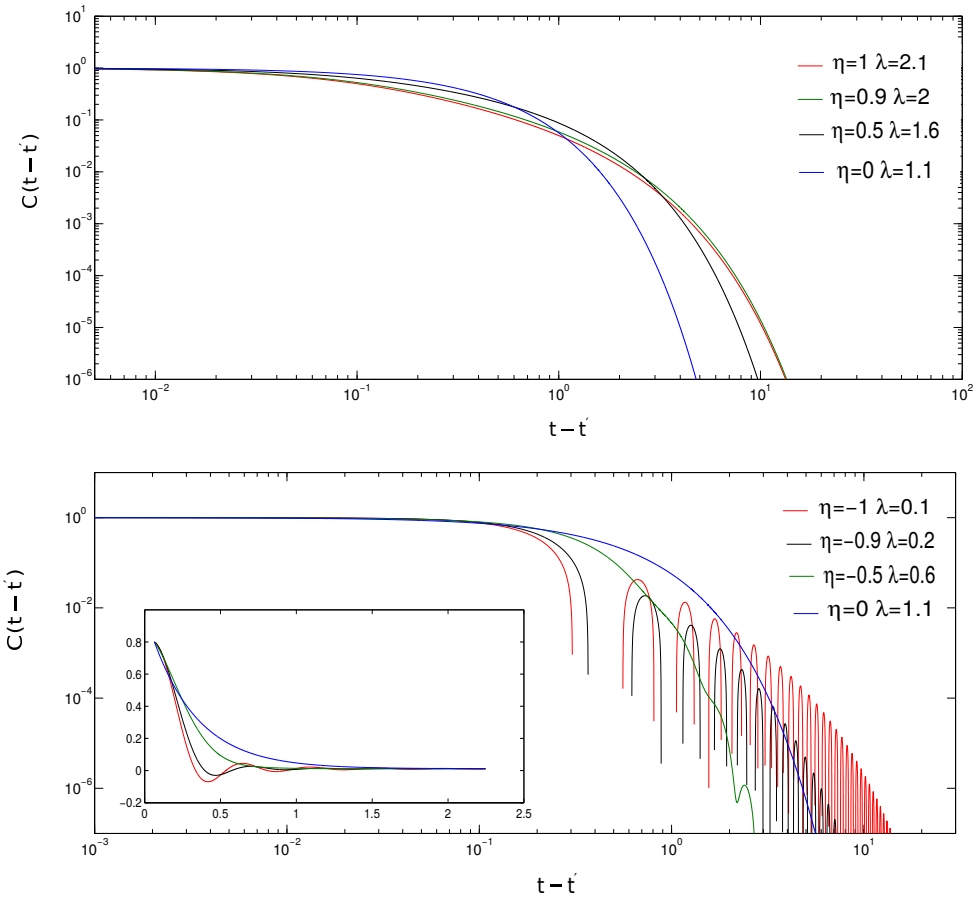

Figure 2.8: Numerical Fourier transform of $\tilde{C}(\omega)$ describing the shape of correlations in temporal domain. (Top) Different positive symmetries: the combination of power laws and an exponential cut-off can be noted. (Bottom) Different negative symmetries: the emergence of oscillations for decreasing η is visible. The normalization is chosen in such a way that $C(t-t) = C(0) = 1$.

As a consequence, the response decays asymptotically as

$$R(t-t') \sim \frac{\mathrm{e}^{-(\lambda-2)(t-t')}}{\sqrt{4\pi}(t-t')^{3/2}} \tag{2.5.24}$$

For the correlation function, if we substitute the expression (2.5.23) into (2.5.19) and carry out the integration we obtain the asymptotic behaviour

$$C(t-t') \sim \frac{1}{2\sqrt{\pi(t-t')}} F((\lambda-2)(t-t')) \tag{2.5.25}$$

where

$$F(x) = \mathrm{e}^{-x} - \sqrt{\pi x}\,\mathrm{erfc}(x) \tag{2.5.26}$$

Two regimes can be distinguished: for $x \ll 1$ (i.e. $t-t' \ll 1/(\lambda-2)$) $F(x) \sim 1$ and one has $C(t-t') \sim 1/2\sqrt{\pi(t-t')}$, whereas for $x \gg 1$ (i.e. $t-t' \gg 1/(\lambda-2)$) $F(x) \sim \mathrm{e}^{-x}/2x$ thus $C(t-t') \sim \mathrm{e}^{-(\lambda-2)(t-t')}/4\sqrt{\pi}(t-t')^{3/2}(\lambda-2)$. A comparison between the exact (2.5.19) and the asymptotic (2.5.25) expressions for the correlation function is shown in figure 2.9.

If $\boldsymbol{K}$ is anti-symmetric ($\eta = -1$), one can perform largely analogous calculations. The explicit expression for the correlations is

$$C(t,t') = \Sigma \int_0^{\min(t,t')} dt''\, \mathrm{Tr}\,\mathrm{e}^{-\lambda(t+t'-2t'')+\boldsymbol{K}(t-t')} = \frac{\Sigma}{2\lambda}\Big(\mathrm{e}^{-\lambda|t-t'|} - \mathrm{e}^{-\lambda(t+t')}\Big)\mathrm{Tr}\,\mathrm{e}^{\boldsymbol{K}(t-t')} \tag{2.5.27}$$

Replacing the trace by an integral over the eigenvalue spectrum, which is now a Wigner semicircle rotated onto the imaginary axis, and taking the long-time limit gives the TTI form

$$C(t-t') = \frac{\Sigma}{2\lambda}\frac{J_1(2(t-t'))}{t-t'}\mathrm{e}^{-\lambda(t-t')} \tag{2.5.28}$$

The Bessel function of the first kind in this is related to the modified Bessel function by $J_1(\mathrm{i}x) = iI_1(x)$ (see e.g. [47]). The response function for $t > t'$ is found similarly as

$$R(t-t') = \theta(t-t')\frac{J_1(2(t-t'))}{t-t'}\mathrm{e}^{-\lambda(t-t')} \tag{2.5.29}$$

From the asymptotics of J_1 one then finds for large time differences

$$C(t-t') \sim \frac{\mathrm{e}^{-\lambda(t-t')}\,\sin\left[2((t-t')-\frac{\pi}{8})\right]}{\lambda\sqrt{\pi}(t-t')^{3/2}} \tag{2.5.30}$$

so the power law component of the decay is as for the symmetric case $\eta = 1$, but here with an oscillatory modulation from the exponential. For a comparison between the exact (2.5.28) and the asymptotic (2.5.30) expressions for the correlation we refer to figure 2.10.

The results (2.5.28) and (2.5.29) show that correlation and response are fully proportional for $\eta = -1$. This is unexpected from the point of view of the FDT, but of course here we are considering

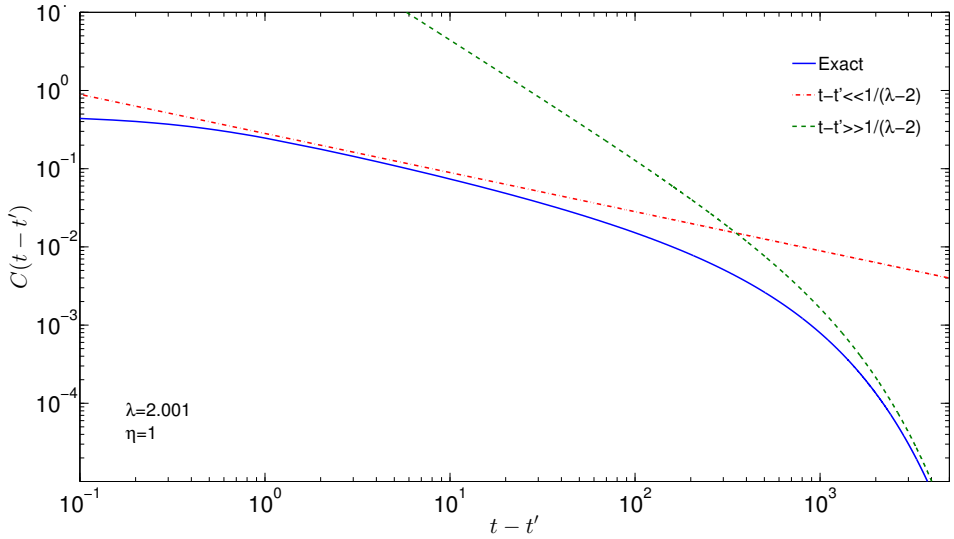

Figure 2.9: Correlations in the time domain: comparison between the analytically exact expression and the asymptotic curves for small and large $t - t'$, for symmetric interactions, $\eta = 1$.

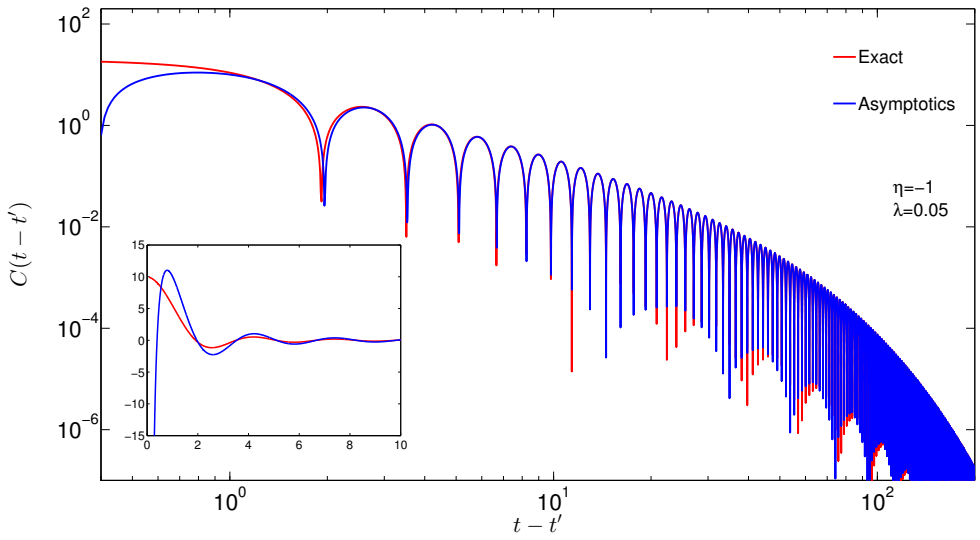

Figure 2.10: Analogue of figure 2.9 for antisymmetric interactions, $\eta = -1$. The power law decay is visible here in the envelope of the oscillatory relaxation.

interactions that are not symmetric. Probability currents are then generically present in the steady state. These translate into additional terms in the FDT, giving rise to a Modified Fluctuation-Dissipation Theorem (MFDT); see e.g. [48]. One can check that these terms generate exactly the proportionality between correlation and response we found above. Let us investigate this aspect closer.

Let us assume $x_i(t)$ as the main observable of the system, i.e. the one for which we consider correlations and responses. Recall the vectorial form of the generic dynamics we are analyzing

$$\frac{d\boldsymbol{x}(t)}{dt} = -\lambda \boldsymbol{x}(t) + \boldsymbol{K}\boldsymbol{x}(t) + \boldsymbol{\xi}(t) \tag{2.5.31}$$

The statistics over dynamical trajectories is Gaussian, thus we can write the probability density $\rho(t)$ as

$$\rho(t) \sim \mathrm{e}^{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{C}^{-1}\boldsymbol{x}} \tag{2.5.32}$$

where $\boldsymbol{C}$ is the steady state correlation (it must be intended as $\boldsymbol{C}(0) = \boldsymbol{C}(t-t)$, an equal time correlator in the TTI regime) and it satisfies the Lyapunov equation

$$(-\lambda + \boldsymbol{K})^T \boldsymbol{C} + \boldsymbol{C}(-\lambda + \boldsymbol{K}) + \boldsymbol{\Sigma} = 0 \tag{2.5.33}$$

The current $\boldsymbol{j}(t)$ related to the probability density $\rho(t)$ is given by

$$\boldsymbol{j}(t) = \left[(-\lambda + \boldsymbol{K})\boldsymbol{x}(t) - \frac{\Sigma}{2}\nabla\right]\rho(t) = \left[(-\lambda + \boldsymbol{K})\boldsymbol{x}(t) + \frac{\Sigma}{2}\boldsymbol{C}^{-1}\boldsymbol{x}(t)\right]\rho(t) \tag{2.5.34}$$

The presence of non-conservative forces in the dynamics can be formalized by defining an induced observable

$$B_i(t) = \rho(t)^{-1}\boldsymbol{j}(t)\cdot\nabla x_i(t) = \left[\left(-\lambda + \boldsymbol{K} + \frac{\Sigma}{2}\boldsymbol{C}^{-1}\right)\boldsymbol{x}(t)\right]_i \tag{2.5.35}$$

We can then define

$$\mathcal{B}_{ij}(t-t') = \theta(t-t')\langle x_i(t)B_j(t')\rangle \tag{2.5.36}$$

The MFDT [48] for $t > t'$ would then read as follows

$$\frac{\Sigma}{2}R_{ij}(t-t') = \frac{\partial}{\partial t'}C_{ij}(t-t') - \mathcal{B}_{ij}(t-t') \qquad \forall \quad i,j \tag{2.5.37}$$

or in matrix form

$$\frac{\Sigma}{2}\boldsymbol{R}(t-t') = \frac{\partial}{\partial t'}\boldsymbol{C}(t-t') - \boldsymbol{\mathcal{B}}(t-t') \tag{2.5.38}$$

Note that we are looking at the thermodynamic limit, in which an average over the disorder is performed. Thus we take the trace of these matrices

$$\frac{\Sigma}{2}\mathrm{Tr}\,\boldsymbol{R}(t-t') = \frac{\partial}{\partial t'}\mathrm{Tr}\,\boldsymbol{C}(t-t') - \mathrm{Tr}\,\boldsymbol{\mathcal{B}}(t-t') \tag{2.5.39}$$

and in particular

$$\begin{aligned}\operatorname{Tr}\boldsymbol{\mathcal{B}}(t-t') &= \sum_i \mathcal{B}_{ii}(t-t') = \sum_i \left\langle x_i(t)\left[\Big(-\lambda+\boldsymbol{K}+\frac{\Sigma}{2}\boldsymbol{C}^{-1}\Big)\boldsymbol{x}(t')\right]_i\right\rangle = \\ &= \sum_{ij}\Big(-\lambda+\boldsymbol{K}+\frac{\Sigma}{2}\boldsymbol{C}^{-1}\Big)_{ij}\langle x_j(t')x_i(t)\rangle = \\ &= \operatorname{Tr}\left[\Big(-\lambda+\boldsymbol{K}+\frac{\Sigma}{2}\boldsymbol{C}^{-1}\Big)\boldsymbol{C}(t-t')\right] \end{aligned} \tag{2.5.40}$$

If we assume that $\boldsymbol{K}$ is diagonalizable, thus normal ($\boldsymbol{K}\boldsymbol{K}^T = \boldsymbol{K}^T\boldsymbol{K}$), the following explicit form for $\boldsymbol{C}$ can be found

$$\boldsymbol{C} = \int_0^{+\infty} dt\, \mathrm{e}^{(-\lambda+\boldsymbol{K})^T t}\,\boldsymbol{\Sigma}\,\mathrm{e}^{(-\lambda+\boldsymbol{K})t} = \Sigma\int_0^{+\infty} dt\,\mathrm{e}^{-2\lambda t}\,\mathrm{e}^{(\boldsymbol{K}+\boldsymbol{K}^T)t} = -\Sigma(-2\lambda+\boldsymbol{K}+\boldsymbol{K}^T)^{-1} \tag{2.5.41}$$

where we remark that $\boldsymbol{\Sigma} = \Sigma\mathbb{1}$. As a consequence

$$\boldsymbol{C}^{-1} = -\frac{1}{\Sigma}(-2\lambda+\boldsymbol{K}+\boldsymbol{K}^T) \tag{2.5.42}$$

and

$$\mathcal{B}(t-t') = \operatorname{Tr}\boldsymbol{\mathcal{B}}(t-t') = \operatorname{Tr}\Big[\frac{1}{2}(\boldsymbol{K}-\boldsymbol{K}^T)\boldsymbol{C}(t-t')\Big] \tag{2.5.43}$$

The additional term is connected to the antisymmetric part of $\boldsymbol{K}$: if $\boldsymbol{K}$ is symmetric, it vanishes and the ordinary FDT is recovered. If $\boldsymbol{K}$ is antisymmetric, we can write

$$\begin{aligned}\mathcal{B}(t-t') &= \operatorname{Tr}\Big[\frac{1}{2}(\boldsymbol{K}-\boldsymbol{K}^T)\boldsymbol{C}(t-t')\Big] = \int_{-2}^{2}\frac{d\tilde{k}}{2\pi}\sqrt{4-\tilde{k}^2}\,\mathrm{i}\tilde{k}\,C_k(t-t') = \\ &= \Sigma\int_{-2}^{2}\frac{d\tilde{k}}{2\pi}\sqrt{4-\tilde{k}^2}\,\mathrm{i}\tilde{k}\,\mathrm{e}^{-\mathrm{i}\tilde{k}(t-t')}\int_{-\infty}^{t'} dt_1\mathrm{e}^{-\lambda(t+t'-2t_1)} = \\ &= \frac{\Sigma}{2\lambda}\frac{d}{dt'}\Big(\frac{J_1(2(t-t'))}{(t-t')}\Big)\mathrm{e}^{-\lambda(t-t')} \end{aligned} \tag{2.5.44}$$

where

$$\begin{aligned}&\textstyle\int_{-\infty}^{t'} dt_1\mathrm{e}^{-\lambda(t+t'-2t_1)} = \frac{\mathrm{e}^{-\lambda(t-t')}}{2\lambda} \\ &\textstyle\int_{-2}^{2}\frac{d\tilde{k}}{2\pi}\sqrt{4-\tilde{k}^2}\,\mathrm{i}\tilde{k}\,\mathrm{e}^{-\mathrm{i}\tilde{k}(t-t')} = \frac{d}{dt'}\Big(\int_{-2}^{2}\frac{d\tilde{k}}{2\pi}\sqrt{4-\tilde{k}^2}\,\mathrm{e}^{-\mathrm{i}\tilde{k}(t-t')}\Big) = \frac{d}{dt'}\Big(\frac{J_1(2(t-t'))}{(t-t')}\Big)\end{aligned}$$

The MFDT (2.5.39) for the average correlations (2.5.28) and responses (2.5.29) can be verified

$$\begin{aligned}\frac{\Sigma}{2}R(t-t') &= \frac{\Sigma}{2}\frac{J_1(2(t-t'))}{(t-t')}\mathrm{e}^{-\lambda(t-t')} = \\ &= \frac{d}{dt'}\Big(\frac{\Sigma}{2\lambda}\frac{J_1(2(t-t'))}{(t-t')}\mathrm{e}^{-\lambda(t-t')}\Big) - \frac{\Sigma}{2\lambda}\frac{d}{dt'}\Big(\frac{J_1(2(t-t'))}{(t-t')}\Big)\mathrm{e}^{-\lambda(t-t')} = \\ &= \frac{d}{dt'}C(t-t') - \mathcal{B}(t-t') \end{aligned} \tag{2.5.45}$$

2.6 Discussion and Conclusion

In this chapter we have developed and studied a novel approach for deriving approximate descriptions for large dynamical systems with continuous degrees of freedom. We refer to the method as an “extended Plefka expansion", where the extension lies in including second order statistics of the fluctuating degrees of freedom in the set of order parameters, rather than only first order averages, i.e. means. Expanding in second order of interaction strength, we derive from the original dynamics – a system of coupled stochastic differential equations – effective equations of motion for each single degree of freedom. These equations are decoupled, with interactions being represented by an effective noise that is no longer white, and a memory term that connects each degree of freedom to its own past. The parameters governing these effective interaction terms are obtained from deterministic (nonlinear) coupled equations.

One key question we studied is under what circumstances the extended Plefka expansion can give exact results for large systems. We demonstrated explicitly for a linear dynamical model that this exactness holds when couplings are of mean field type, i.e. weak and long-ranged. An analogy can be drawn with works on soft spins dynamics [49, 50], where the exact infinite-range limit produces local mean field equations with self-consistent propagator and noise. Importantly, the agreement we show holds independently of whether the dynamics obeys detailed balance, due to symmetry in the interaction coefficients, or not; we explored the entire range of symmetry parameters from symmetry ($\eta = 1$) to asymmetry ($\eta = 0$) to anti-symmetry ($\eta = -1$). We also studied the quantitative features of the model in some detail, focussing on correlation functions and power spectra as their Fourier transform; this analysis revealed non-trivial crossover phenomena in the vicinity of either full symmetry or full anti-symmetry.

The extended Plefka method makes exact predictions for our linear model system, whereas – as we discussed – a conventional Plefka expansion fails to predict any non-trivial effects in correlations and responses. This suggests our method as a promising candidate for the accurate reconstruction of the dynamics of large systems also in generic nonlinear settings that cannot be solved analytically. The equations we have derived can be applied directly to such a generic case. We have merely restricted ourselves to a model without self-interactions beyond the basic linear one that we assume, but this restriction can easily be lifted at the expense of longer expressions for the memory functions and effective noise correlations (we refer to appendix C for details).

An important question for such future applications is in what other scenarios one would expect the extended Plefka method to become exact in the large system limit. Generalizing from our

linear model one could consider e.g. nonlinear drift terms of the form $\phi_i(\boldsymbol{x}) = \sum_j K_{ij} g(x_j)$, where $g(\boldsymbol{x})$ is a generic non-linear function. With the potential application in biochemical networks in mind, $K_{ij} g(x_j)$ could describe the interaction due to reactions between different species i and j, K_{ij} being reaction rates. From central limit theorem arguments one would expect that the dynamics of such a nonlinear system would again be described exactly by the extended Plefka method, provided that the K_{ij} are weak and long-ranged. This should hold even if the nonlinearities are made species-specific so that $g(x_j)$ is replaced by $g_j(x_j)$. Related models can be found in the context of neural networks, where $g(x_j)$ plays the role of a nonlinear gain function combining "inputs" to determine certain "outputs". The mean field properties of such models, in the case of asymmetric K_{ij} and $g(x_j)$ of sigmoid shape, were studied by Sompolinsky and coworkers [51] and are consistent with the extended Plefka predictions. In general one could think of other simple scenarios where some moments of the variables x_i can be calculated exactly and these may also provide useful future testbeds for our method. Interestingly, after the completion of this work, we discovered that an alternative perturbative approach also taking into account second moments had already been applied by Biroli in the derivation of dynamical TAP equations for the p-spin spherical model [52]. These TAP equations are the fixed-disorder analogue of the disorder-averaged equations first derived by [53]. We have checked that the extended Plefka expansion gives back exactly Biroli's equations when applied to the p-spin model, with the Lagrange multiplier for the spherical constraint playing the role of our λ_i: see appendix C. This is an important consistency check. Nevertheless we stress that the framework discussed in this chapter is in principle wider, encompassing generic continuous variables and generic nonlinear interaction terms. In addition, it is aimed at producing approximate decoupled equations that could be regarded as the first step for implementing *inference* algorithms.

A promising further development of our method would be to find a more sophisticated treatment of nonlinear self-interactions. In our present approach, these would be subsumed into the general interaction terms. Alternatively one could try to treat nonlinear self-interactions exactly, by keeping them as part of the non-interacting baseline for the Plefka expansion. This would result in effective equations of motion that are still decoupled but now nonlinear and driven by memory terms and coloured noise. The resulting self-consistency conditions for the order parameters would then have to be obtained by simulation, but there are precedents [39] for doing this in a computationally efficient manner.

A further direction for future work would be to understand in more detail the relation to the Expectation-Propagation (EP) algorithm [54, 55]. For the case of linear self-interactions $-\lambda_i x_i$

that we mostly focussed on, EP and our extended Plefka method both yield factorized (over degrees of freedom) probability distributions over system trajectories, with the same non-interacting Gaussian baseline. It would therefore be interesting to clarify what the differences between the approaches are and under what circumstances they might lead to identical approximations.

One important simplification we had to make was to assume that the different degrees of freedom x_i are affected by independent noise, so that the noise covariance matrix $\mathbf{\Sigma}$ is diagonal. On the other hand, in biochemical networks there are generically off-diagonal noise correlations: noise arises from the stochasticity of when reactions take place, and each non-trivial reaction affects the number of molecules from several molecular species. The extension of our approach to this case requires further work. If the noise covariance $\mathbf{\Sigma}$ is at least independent of the state $\boldsymbol{x}$ of the system – though even this is not the generic case for reaction networks – then one could imagine transforming the variables x_i linearly to diagonalize $\mathbf{\Sigma}$. This would then make our approach directly applicable, but would also make the biological interpretation of any predictions rather less intuitive.

In the long term our approximation framework should also help one to tackle network reconstruction problems, and this is a further important direction for future work. In fact, once the forward dynamics has been fully characterized as we have done here, one can think of setting up inverse techniques based on the same description. This would allow one e.g. to infer the states of hidden (unobserved) nodes from observations of other (visible) variables, as we will see in chapter 3, and ultimately to learn interaction parameters and hence network structure from data.

Appendix A

Residue calculation

A.1 Residue calculation

We provide some details here of the exact calculation of the Laplace transformed correlation function $\tilde{C}(z)$ for the linear model with weak long-range interactions. The singularity structure of this function in the complex z-plane is sketched in figures A.1, A.2, A.3, A.4 and A.5. Below it will be useful to remember also that the singularities of the response function $\tilde{R}(z)$ are the same as those singularities of $\tilde{C}(z)$ that lie in the left half-plane. The difference arises because in the time domain the correlation function $C(t-t')$ is even in $t-t'$, while the response $R(t-t')$ vanishes for $t-t' < 0$.

Let us begin with asymmetric random interactions ($\eta = 0$), for which (2.4.27) can be decomposed as

$$\mathrm{Tr}\,\tilde{\boldsymbol{C}}(z) = \Sigma(I_1 + I_2) \tag{A.1.1}$$

where

$$I_1 = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{1}{z - \mathrm{i}\omega} \frac{1}{1 - (\lambda + z)(\lambda - \mathrm{i}\omega)} \tag{A.1.2}$$

$$I_2 = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{1}{z - \mathrm{i}\omega} \frac{1}{(\lambda - z)(\lambda + \mathrm{i}\omega) - 1} \tag{A.1.3}$$

These integrals can be performed in the complex plane as parts of integrals along a closed path. In fact, if we denote the integrands as $f_{1,2}(\omega)$, we can write

$$I_{1,2} = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} f_{1,2}(\omega) = \Big(\int_{-\infty}^{+\infty} + \int_C \Big) \frac{d\omega}{2\pi} f_{1,2}(\omega) = \oint \frac{d\omega}{2\pi} f_{1,2}(\omega) = \frac{1}{2\pi}\, 2\pi\mathrm{i} \sum_i \mathrm{Res} f_{1,2}(\omega)|_{\omega=\omega_i} \tag{A.1.4}$$

Here the ω_i refer to the poles inside the closed path, as drawn in figure A.6. The value of the integral along the semicircle C vanishes when the radius goes to infinity as $f_{1,2}(\omega) \approx \frac{1}{\omega^2} \to 0$ for

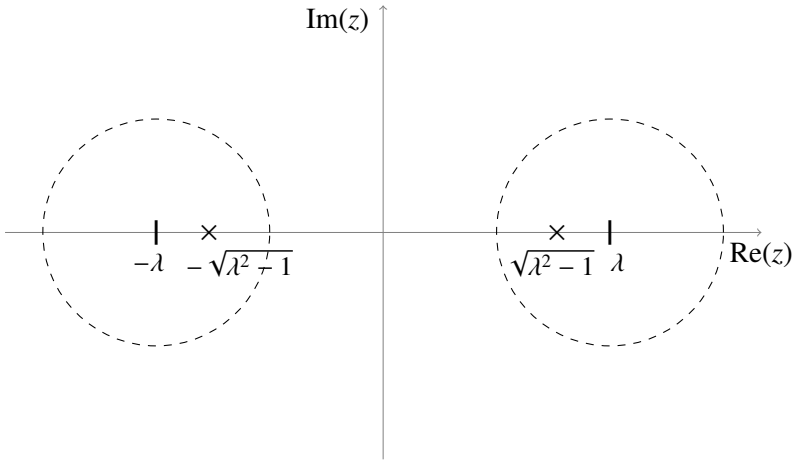

Figure A.1: Singularities of $\tilde{C}(z)$ in the complex z-plane for asymmetric interactions ($\eta = 0$): for this value of η the only singularities are the two poles $z_{\text{pole}} = \pm\sqrt{\lambda^2 - 1}$. The random matrix calculation following [44], which uses a perturbative approach, applies only outside the two copies of the spectral circle (of radius one) shifted to be centred at $z = \pm\lambda$, but the results can be continued analytically into the circles as $\tilde{C}(z)$ is a Laplace transform.

$|\omega| \to \infty$. The poles for $f_1(\omega)$ are

$$\omega_1 = -\mathrm{i}z \qquad \omega_2 = \mathrm{i}\left[\frac{1-\lambda(\lambda+z)}{\lambda+z}\right] \tag{A.1.5}$$

while the poles for $f_2(\omega)$ are

$$\omega_1 = -\mathrm{i}z \qquad \omega_3 = -\mathrm{i}\left[\frac{1-\lambda(\lambda-z)}{\lambda-z}\right] \tag{A.1.6}$$

To locate the poles in the complex ω-plane we can fix a convenient region for the value of z, which for the purposes of our integration is an external parameter, and then continue the result analytically in z at the end. In particular, it is useful to ensure that $z + \lambda$ and $\lambda - z$ are kept outside the support of the spectrum of the interaction matrix $\boldsymbol{K}$, i.e. that z stays outside the circles in figure A.1. Let us therefore choose z as real and $z > \lambda + 1$. With this choice, ω_1 and ω_2 lie on the negative imaginary axis, while ω_3 lies on the positive axis. We thus close the integration contour in the upper half plane (see figure A.6) so that it includes only ω_3 and obtain

$$I_1 + I_2 = 2\pi\mathrm{i}\,\mathrm{Res} f_1(\omega)|_{\omega=\omega_2} = \frac{1}{2\pi}\,2\pi\mathrm{i}\,\frac{1}{\mathrm{i}\,(\lambda^2 - z^2 - 1)} = \frac{1}{\lambda^2 - z^2 - 1} \tag{A.1.7}$$

Thus

$$\tilde{C}(z) = \frac{\Sigma}{\lambda^2 - z^2 - 1} \tag{A.1.8}$$

This exact result agrees with the prediction (2.4.13) of the extended Plefka method once we insert the appropriate expression (2.4.10) for the response in the asymmetric case.

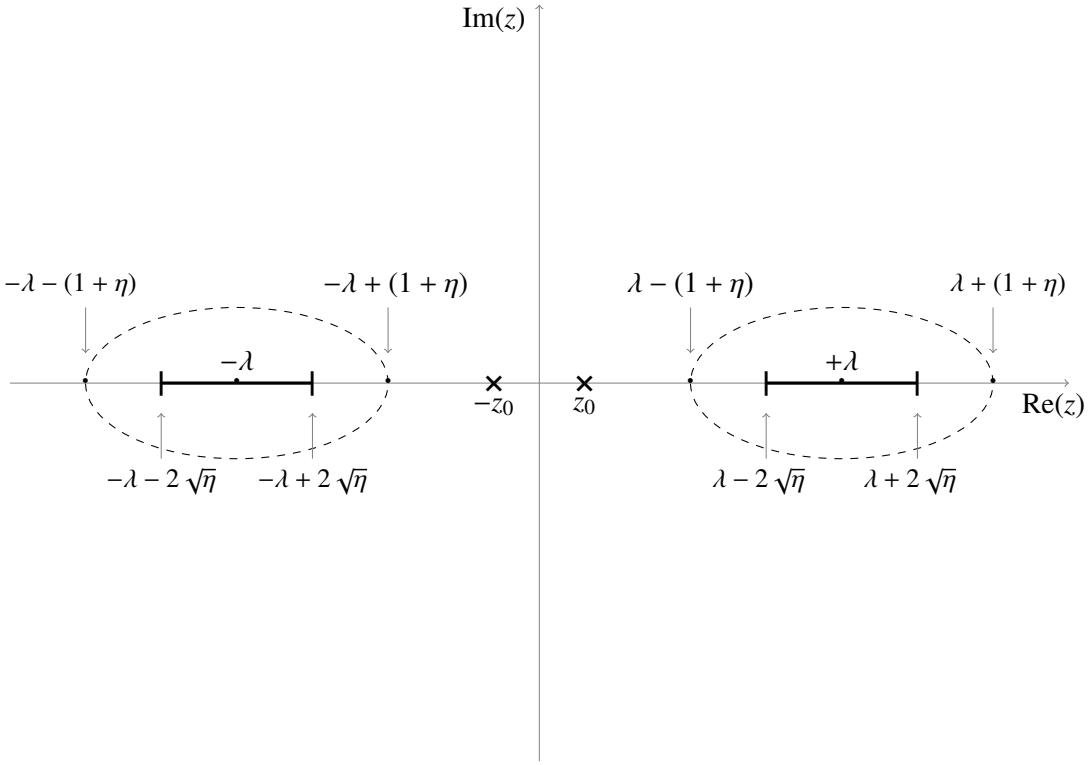

Figure A.2: Singularities of $\tilde{C}(z)$ in the complex z-plane for generic positive interaction symmetry ($\eta > 0$): there are two poles $z_{\text{pole}} = \pm z_0$ as well as two branch cuts connecting the four points $z_{\text{bc}} = \pm\lambda \pm 2\sqrt{\eta}$. The random matrix calculation applies only outside the two copies of the spectral ellipse shifted to be centred at $z = \pm\lambda$. The ellipses have real and imaginary semi-axes $1 + \eta$ and $1 - \eta$, respectively; the foci are the edges of the branch cuts.

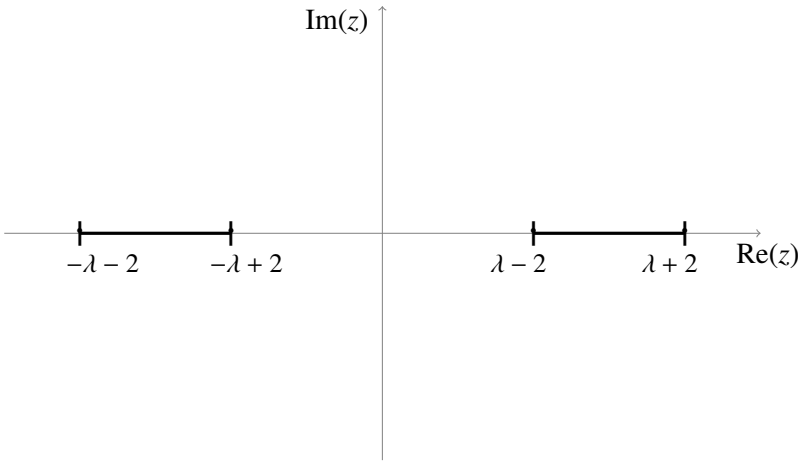

Figure A.3: Region of the complex plane where $\tilde{C}(z)$ for symmetric couplings is defined, as the particular case $\eta = 1$ of figure A.2: branch cuts occur for $-\lambda - 2 < z < -\lambda + 2$ and $\lambda - 2 < z < \lambda + 2$.

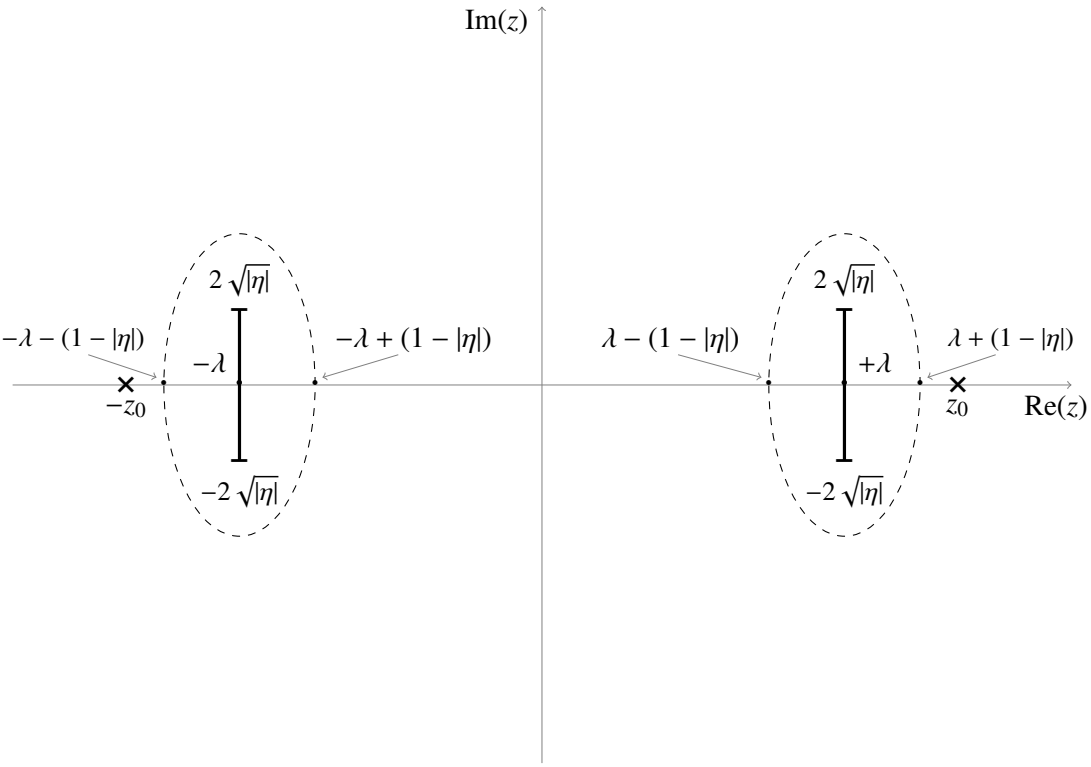

Figure A.4: Analogue of figure A.2 for negative symmetry parameters η. The branch cuts and major semi-axes of the ellipses are now along the imaginary rather than the real direction, so that the foci and hence the edges of the branch cuts are at $z = \pm\lambda \pm 2\mathrm{i}\sqrt{|\eta|}$.

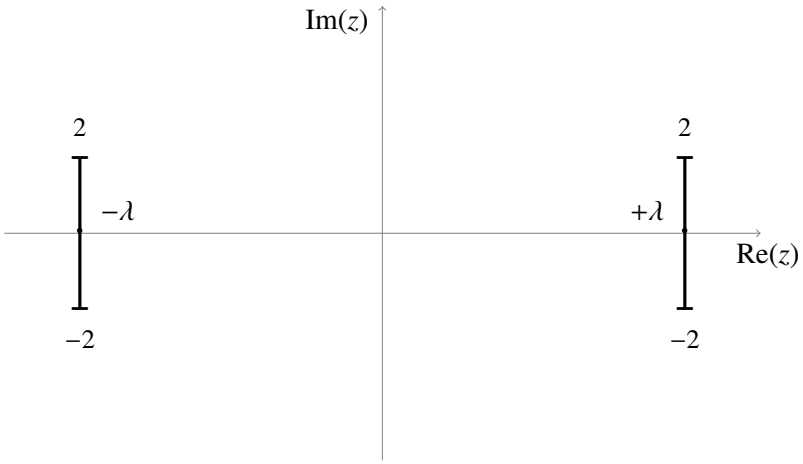

Figure A.5: Region of the complex plane where $\tilde{C}(z)$ for anti-symmetric couplings $\eta = -1$ is defined.

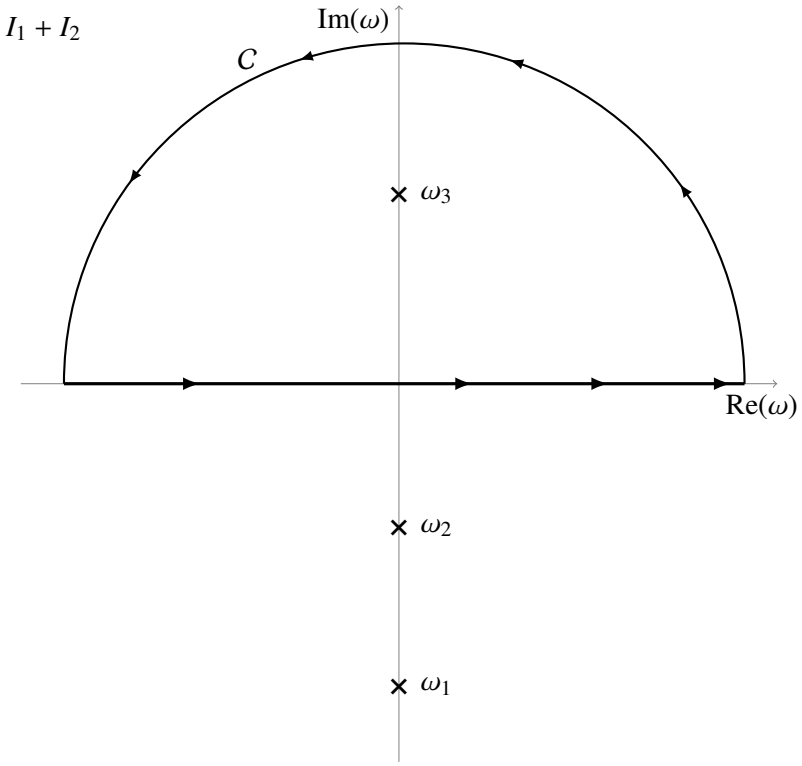

Figure A.6: Integration contour in the complex plane for $I_1 + I_2$ in the asymmetric case.

For arbitrary correlations η between K_{ij} and K_{ij}, (2.4.30) can be rewritten as

$$\begin{aligned}\tilde{C}(z) &= \Sigma \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi}[\tilde{R}(-z)f_1(\omega) + \tilde{R}(z)f_2(\omega)] = \\ &= \Sigma\Big(- \oint_1 \frac{d\omega}{2\pi}\tilde{R}(-z)f_1(\omega) + \oint_2 \frac{d\omega}{2\pi}\tilde{R}(z)f_2(\omega) + \int_{C_1} \frac{d\omega}{2\pi}\tilde{R}(-z)f_1(\omega) - \int_{C_2} \frac{d\omega}{2\pi}\tilde{R}(z)f_2(\omega)\Big) = \\ &= \frac{\Sigma}{2\pi} 2\pi\mathrm{i}\Big(- \sum_i \mathrm{Res}[\tilde{R}(-z)f_1(\omega)]\Big|_{\omega=\omega_i} + \sum_j \mathrm{Res}[\tilde{R}(z)f_2(\omega)]\Big|_{\omega=\omega_j}\Big) \end{aligned} \tag{A.1.9}$$

where the signs refer to integration contours arranged as figures A.7 and A.8, i.e. with an anti-clockwise orientation. The functions f_1 and f_2 are defined as

$$f_1(\omega) = \frac{1}{z - \mathrm{i}\omega} \frac{\frac{\lambda+\mathrm{i}\omega-\sqrt{(\lambda+\mathrm{i}\omega)^2-4\eta}}{2\eta}}{1 - \tilde{R}(-z)\Big(\frac{\lambda+\mathrm{i}\omega-\sqrt{(\lambda+\mathrm{i}\omega)^2-4\eta}}{2\eta}\Big)} = \frac{1}{z - \mathrm{i}\omega} \frac{\tilde{R}(\mathrm{i}\omega)}{1 - \tilde{R}(-z)\tilde{R}(\mathrm{i}\omega)} \tag{A.1.10}$$

$$f_2(\omega) = \frac{1}{z - \mathrm{i}\omega} \frac{\frac{\lambda-\mathrm{i}\omega-\sqrt{(\lambda-\mathrm{i}\omega)^2-4\eta}}{2\eta}}{1 - \tilde{R}(z)\Big(\frac{\lambda-\mathrm{i}\omega-\sqrt{(\lambda-\mathrm{i}\omega)^2-4\eta}}{2\eta}\Big)} = \frac{1}{z - \mathrm{i}\omega} \frac{\tilde{R}(-\mathrm{i}\omega)}{\tilde{R}(z)\tilde{R}(-\mathrm{i}\omega) - 1} \tag{A.1.11}$$

In the last line of (A.1.9) we have already exploited that, because $\lim_{|\omega|\to\infty}|\omega f_1(\omega)| = 0$ and $\lim_{|\omega|\to\infty}|\omega f_2(\omega)| = 0$, the contributions from the semicircles C_1 and C_2 vanish when their radius is sent to infinity.

For further evaluation we first focus on $\eta > 0$. As before it is convenient to restrict z, here such that it lies outside the left spectral ellipses in figure A.2. The denominator of $f_1(\omega)$ and $f_2(\omega)$ then has only one relevant zero

$$\omega_1 = -\mathrm{i}z \tag{A.1.12}$$

This is because with z restricted as above, $|\tilde{R}(-z)| < 1$, for exactly the same reason that $|g_1| < 1$ outside the (unshifted) spectral ellipse as discussed in section 2.4.3. Our choice of contour 1 also guarantees that $|\tilde{R}(\mathrm{i}\omega)| < 1$ because the integration contour avoids the appropriately rotated spectral ellipse that governs $\tilde{R}(\mathrm{i}\omega)$, as shown in figure A.7. Thus the denominator $1-\tilde{R}(-z)\tilde{R}(\mathrm{i}\omega)$ in (A.1.10) can never be zero inside our integration contour. An exactly analogous argument applies to the integration over f_2.

We now further restrict z to be real and positive, such that ω_1 lies in the lower half plane (see figures A.7 and A.8). Only the integration contour for f_1 then encircles any singularities at all, and we obtain from (A.1.9)

$$\begin{aligned} \tilde{C}(z) &= \Sigma \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} [\tilde{R}(-z) f_1(\omega) + \tilde{R}(z) f_2(\omega)] = \\ &= -\frac{\Sigma}{2\pi} 2\pi\mathrm{i}\, \mathrm{Res}[\tilde{R}(-z) f_1(\omega)]\Big|_{\omega=\omega_1} = \frac{\Sigma\, \tilde{R}(z)\tilde{R}(-z)}{1 - \tilde{R}(z)\tilde{R}(-z)} \end{aligned} \tag{A.1.13}$$

as claimed in (2.4.37) in the main text.

For $\eta < 0$ an analogous calculation of the correlation function integral can be performed. Some changes in the relevant regions of the complex plane occur, namely the ellipses bounding the support of the spectrum are rotated (compare figures A.2 and A.4), but the method is the same for $\eta > 0$, and so is the result. Integration contours for $\eta < 0$ are sketched in figures A.9 and A.10.

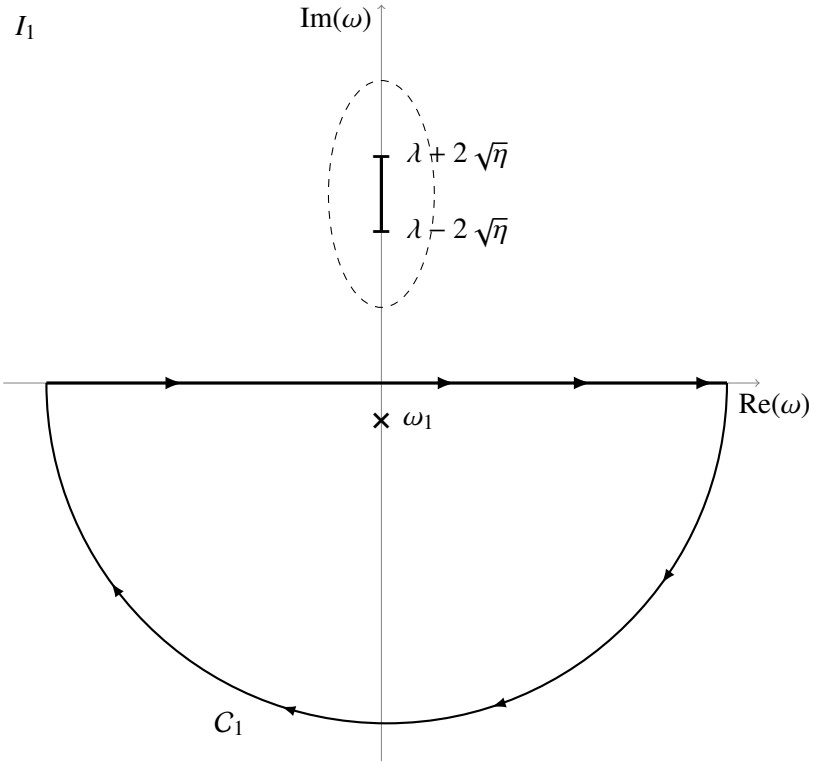

Figure A.7: Integration contour in the complex plane for I_1 in the case of generic symmetry $\eta > 0$.

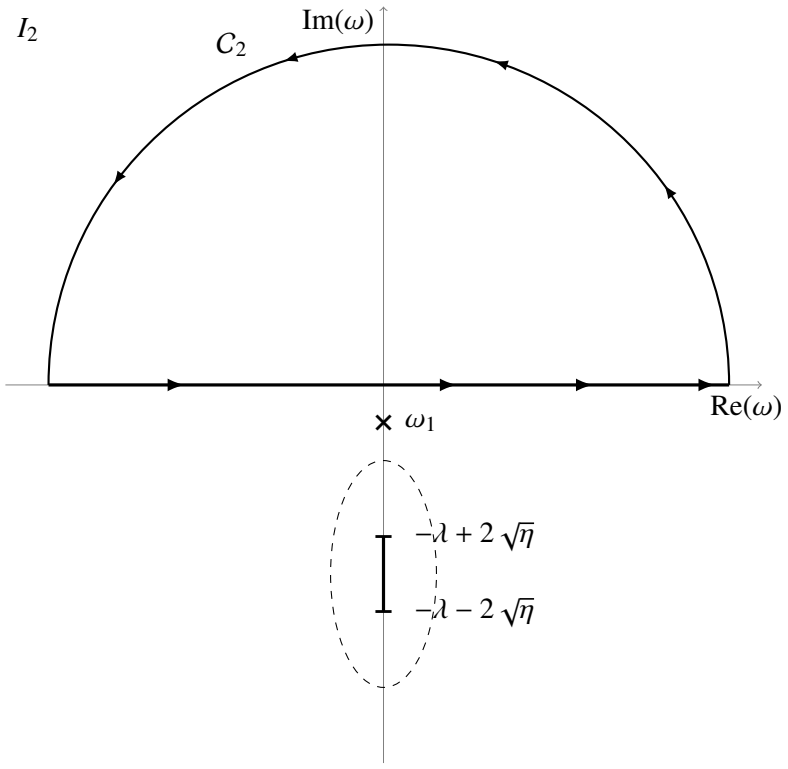

Figure A.8: Integration contour in the complex plane for I_2 in the case of generic symmetry $\eta > 0$.

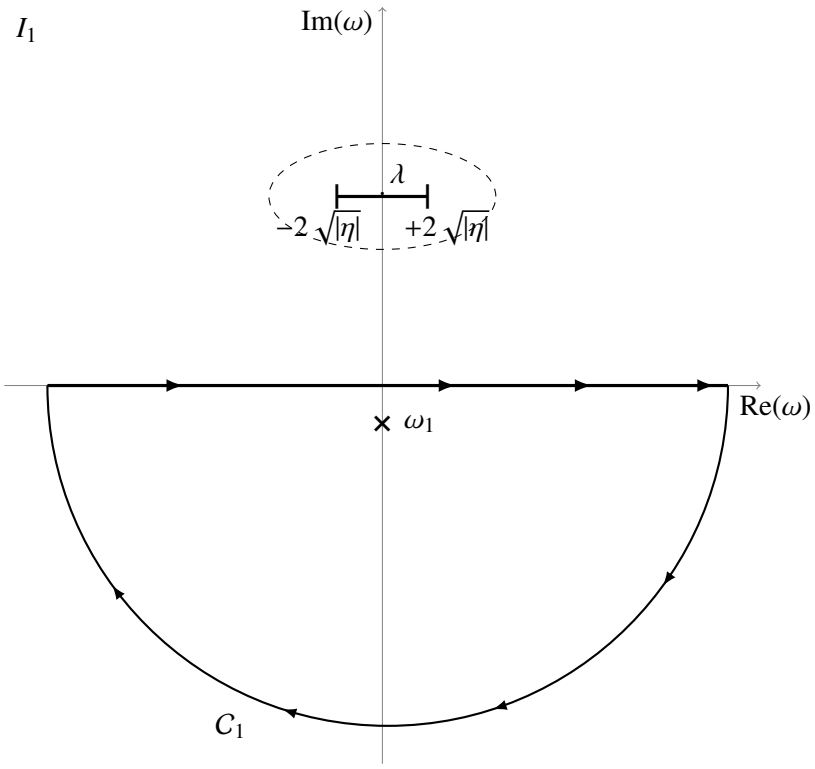

Figure A.9: Integration contour on the complex plane for I_1 in the case of generic symmetry $\eta < 0$.

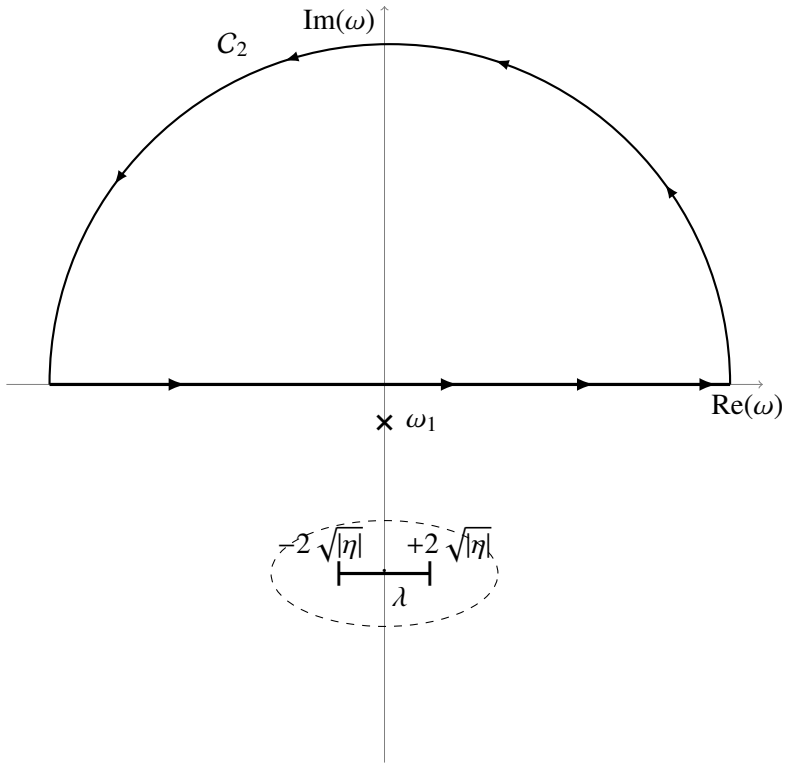

Figure A.10: Integration contour on the complex plane for I_2 in the case of generic symmetry $\eta < 0$.

Appendix B

Inverse Laplace transform

B.1 Inverse Laplace transform

To evaluate the correlations in the time domain, one can take the inverse Laplace transform of expression (2.5.1) in the main text, at least numerically.

Let us set $|t - t'| = \tau$. The inverse double sided Laplace transform of $\tilde{C}(z)$ reads

$$C(\tau) = \frac{1}{2\pi \mathrm{i}} \lim_{T \to +\infty} \int_{-\mathrm{i}T}^{+\mathrm{i}T} \mathrm{e}^{z\tau} \tilde{C}(z) dz \tag{B.1.1}$$

which can be written as the integral along the imaginary axis followed by the integral along C, which vanishes when the radius of the semicircle diverges. Contracting this closed contour then shows that the Laplace Transform is the sum of the integrals taken counterclockwise around the pole (given by the residue) and the branch cut (given by the combination of integrals along c_1, c_2 and along the lines above and below the branch cut, as in figure B.1). In particular one can write

$$\begin{aligned} \oint \mathrm{e}^{z\tau} \tilde{C}(z) dz &= \lim_{T \to +\infty} \int_{-\mathrm{i}T}^{+\mathrm{i}T} \mathrm{e}^{z\tau} \tilde{C}(z) dz + \int_C \mathrm{e}^{z\tau} \tilde{C}(z) dz = 2\pi \mathrm{i}\, \mathrm{Res}\left[\mathrm{e}^{z\tau} \tilde{C}(z)\right]_{z=-z_0} + \int_{c_1} \mathrm{e}^{z\tau} \tilde{C}(z) dz \\ &+ \int_{c_2} \mathrm{e}^{z\tau} \tilde{C}(z) dz - \int_{-\lambda+2\sqrt{\eta}+\mathrm{i}\epsilon_0}^{-\lambda-2\sqrt{\eta}+\mathrm{i}\epsilon_0} \mathrm{e}^{z\tau} \tilde{C}(z) dz - \int_{-\lambda+2\sqrt{\eta}-\mathrm{i}\epsilon_0}^{-\lambda-2\sqrt{\eta}-\mathrm{i}\epsilon_0} \mathrm{e}^{z\tau} \tilde{C}(z) dz \end{aligned} \tag{B.1.2}$$

where ϵ_0 is the radius of c_1 and c_2 semicircles. In the limit of $\epsilon_0 \to 0$

$$\int_{c_1} \mathrm{e}^{z\tau} \tilde{C}(z) dz \to 0 \qquad \int_{c_2} \mathrm{e}^{z\tau} \tilde{C}(z) dz \to 0 \tag{B.1.3}$$

and the integrals on the lines connecting c_1 and c_2 actually run along branch cuts, on the real axis. The radius of C is meant to diverge, in a way that allows the application of results from complex analysis, i.e. the Jordan's lemma

$$\int_C \mathrm{e}^{z\tau} \tilde{C}(z) dz \to 0 \tag{B.1.4}$$

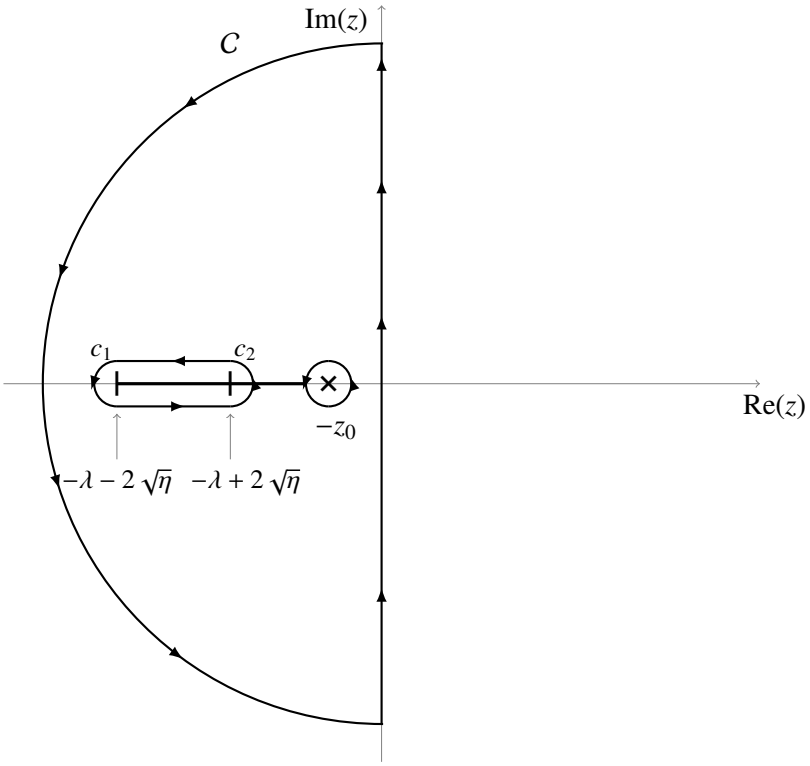

Figure B.1: Integration contour in the complex plane for the inverse double sided Laplace transform of $\tilde{C}(z)$.

Thus one has

$$C(\tau) = \mathrm{Res}[\mathrm{e}^{z\tau}\tilde{C}(z)]_{z=-z_0} - \frac{1}{2\pi\mathrm{i}} \lim_{\epsilon_0 \to 0} \left(\int_{-\lambda-2\sqrt{\eta}+\mathrm{i}\epsilon_0}^{-\lambda+2\sqrt{\eta}+\mathrm{i}\epsilon_0} \mathrm{e}^{z\tau}\tilde{C}(z)dz + \int_{-\lambda+2\sqrt{\eta}-\mathrm{i}\epsilon_0}^{-\lambda-2\sqrt{\eta}+\mathrm{i}\epsilon_0} \mathrm{e}^{z\tau}\tilde{C}(z)dz \right) \quad \text{(B.1.5)}$$

By applying this formula for $\epsilon \to 0$ (i.e. close to symmetry, as $\epsilon = 1 - \eta$) and $\lambda \to \lambda_{\min}$, where $\lambda_{\min} = 1+\eta$, one can verify that in the time domain, at a timescale of order ϵ^{-2}, there is a crossover: the correlation function decay changes from the $\tau^{-1/2}$ form one finds for $\eta = 1$ to an exponential form of decay $\sim \exp(-z_{\mathrm{pole}}\tau)$ (where $z_{\mathrm{pole}} = z_0$). Let us look at this derivation more closely.

First we notice that $\mathrm{e}^{z\tau}\tilde{C}(z)$ can be rewritten as the ratio between two functions

$$\mathrm{e}^{z\tau}\tilde{C}(z) = \frac{f(z)}{d(z)} \quad \text{(B.1.6)}$$

with $f(z) = 4\Sigma\,\mathrm{e}^{z\tau}$ and $d(z) = [(\lambda + z) + \sqrt{(\lambda+z)^2 - 4\eta}][(\lambda - z) + \sqrt{(\lambda-z)^2 - 4\eta}] - 4$, also $d(-z_0) = 0$ by definition. Thus

$$\mathrm{Res}[\mathrm{e}^{z\tau}\tilde{C}(z)]_{z=-z_0} = \left.\frac{f(z)}{d'(z)}\right|_{z=-z_0} = A_{\eta,\lambda}\mathrm{e}^{-z_0\tau} \quad \text{(B.1.7)}$$

The residue contributes with an exponential decay whose amplitude depends on η and λ (through

the expression $z_0 = \left(\frac{1-\eta}{1+\eta}\right)\sqrt{\lambda^2-(1+\eta)^2}$, see (2.5.6)) and the amplitude $A_{\eta,\lambda}$ is given by

$$A_{\eta,\lambda} = \frac{4\Sigma}{\left(\lambda + z_0 + \sqrt{(\lambda+z_0)^2-4\eta}\right)\left(1+\frac{\lambda-z_0}{\sqrt{(\lambda-z_0)^2-4\eta}}\right) + \left(\lambda - z_0 + \sqrt{(\lambda-z_0)^2-4\eta}\right)\left(-1-\frac{\lambda+z_0}{\sqrt{(\lambda+z_0)^2-4\eta}}\right)} \tag{B.1.8}$$

Let us rewrite $\lambda = \lambda_{\min} + \delta\lambda$. In the limit $\lambda \to \lambda_{\min}$ ($\delta\lambda \to 0$), one can verify that

$$A_{\eta,\lambda} \sim z_0^{-1} \sim (\delta\lambda)^{-1/2} \qquad \lim_{\delta\lambda\to 0} A_{\eta,\lambda} = \infty \tag{B.1.9}$$

i.e. this amplitude diverges for $\lambda = \lambda_{\min}$, consistently with the infinite value that we expect for $C(0)$ (from the non-integrable power spectrum $\sim 1/\omega^2$). The limit (B.1.9) makes also the correlator $C(\tau)$ diverge for any τ (including $\tau = 0$). Therefore, also for λ different from $\lambda_{\min}$ but close to it, the residue is the dominating contribution. We can verify this by subtracting from the power spectrum $\tilde{C}(\omega)$ the contribution of the pole expressed in terms of frequencies ω - we shall call it $g_{\text{pole}}(\omega)$

$$g_{\text{pole}}(\omega) = A_{\eta,\lambda}\left(\frac{1}{z_0+z} + \frac{1}{z_0-z}\right)\Big|_{z=\mathrm{i}\omega} = \frac{2A_{\eta,\lambda}z_0}{z_0^2+\omega^2} \tag{B.1.10}$$

This subtraction actually removes the $1/\omega^2$ divergence, as shown in figure B.2 (where we plotted what remains once we take off $g_{\text{pole}}(\omega)$ from the power spectrum) and as can be checked from the $g_{\text{pole}}(\omega)$ limit when $\lambda \to \lambda_{\min}$

$$\lim_{\delta\lambda\to 0} \frac{2A_{\eta,\lambda}z_0}{z_0^2+\omega^2} = \frac{\epsilon^3}{2\omega^2} \tag{B.1.11}$$

precisely the asymptotics given by (2.5.15a). The other contributions to $C(\tau)$ behave as $\tau^{-1/2}$ for $\tau \ll 1/\epsilon^2$ (corresponding to the $1/\sqrt{2\omega}$ law derived in section 2.5.3 of the main text) and there is thus a roughly exponential cutoff for $\tau \gg 1/\epsilon^2$ (only roughly exponential because the master curve for $\omega \sim \epsilon^2$ (2.5.14) is not just a simple Lorentzian). One can finally calculate numerically the integrals in (B.1.5); we refer to figure B.3, where the crossover between these two different regimes is visible.

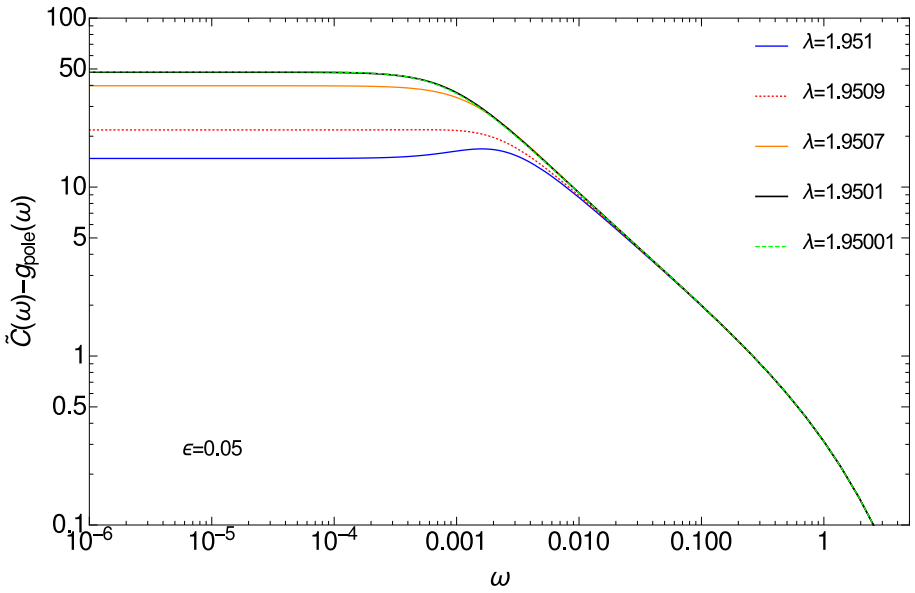

Figure B.2: Subtraction of $g_{\text{pole}}(\omega)$ from the power spectrum $\tilde{C}(\omega)$ for $\epsilon = 0.05$ at varying λ close to λ_{min}. The subtracted spectrum no longer has the $1/\omega^2$ divergence - for $\omega < \epsilon^2$ it levels off to a plateau. For $\delta\lambda \to 0$ the curves tend to overlap as $\tilde{C}(\omega)$ approaches the limit $\epsilon^3/2\omega^2$. For decreasing $\delta\lambda$ one sees that the right boundary of the plateau occurs at smaller ω, i.e. the exponential cutoff kicks in at larger times: in fact it is set by $z_0 \sim \delta\lambda^{1/2}$. The plot is in logarithmic scale and for simplicity $\Sigma = 1$.

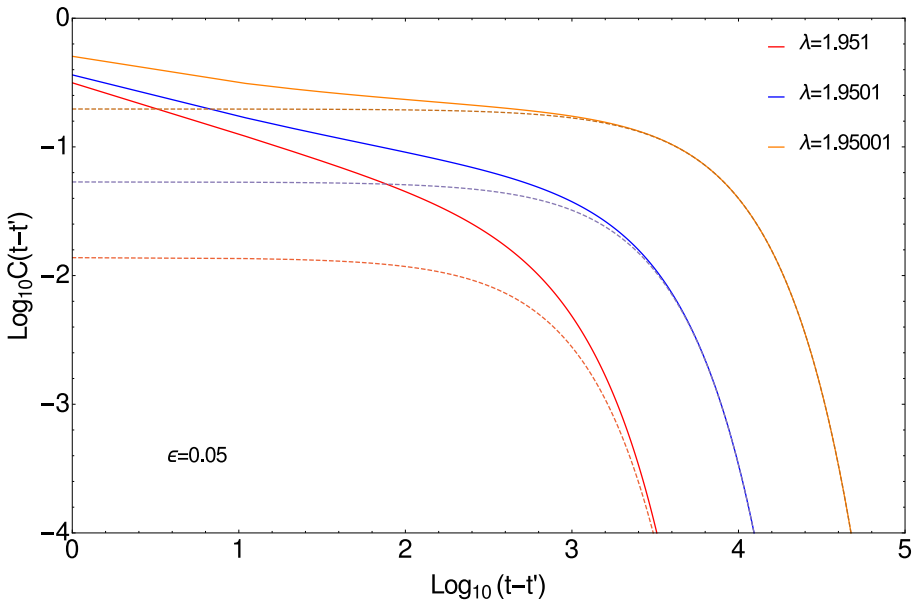

Figure B.3: Numerical inverse Laplace transform for different values of λ at fixed $\epsilon = 0.05$. Values of λ are chosen in such a way that the pole exists. For $t - t' \gg \epsilon^{-2}$, one can observe that the dominating contribution to the curve is given by the residue (dashed lines); the exponential cutoff, determined by z_0, is shifted for different values of λ, as expected. For $t - t' \ll \epsilon^{-2}$ the 3 curves tend to collapse as the power law behaviour $1/\sqrt{2\omega}$, independent of λ, emerges. For simplicity $\Sigma = 1$.

Appendix C

Complete TAP equations

C.1 Complete TAP equations

We lift the restriction $\partial\phi_i(\boldsymbol{x}(t))/\partial x_i(t) = 0$ and the one regarding the additivity of variables in the drift $\phi_i(\boldsymbol{x}(t))$. The second order of the Plefka free energy can be evaluated starting from the equality (2.2.19)

$$G^2 = \left.\frac{\partial^2 G_\alpha}{\partial\alpha^2}\right|_{\alpha=0} = \left\langle\left(\delta\,\frac{d\Xi_\alpha}{d\alpha}\right)^2\right\rangle_0 \tag{C.1.1}$$

Including the first order fields in the effective action, with the prefactor α, and taking $\frac{d\Xi_\alpha}{d\alpha}$ at $\alpha = 0$ gives

$$\begin{aligned}\left.\frac{d\Xi_\alpha}{d\alpha}\right|_{\alpha=0} = &-\Delta\sum_{it}\mathrm{i}\hat{x}_i(t)\phi_i(\boldsymbol{x}(t)) + \Delta\sum_{it}\psi_i^1(t)\Big(x_i(t)-\mu_i(t)\Big) - \Delta\sum_{it}l_i^1(t)\Big(\mathrm{i}\hat{x}_i(t)-\mathrm{i}\hat{\mu}_i(t)\Big) + \\ &+\Delta^2\sum_{itt'}\hat{C}_i^1(t,t')\Big(x_i(t)x_i(t')-C_i(t,t')\Big) + \Delta^2\sum_{itt'}\hat{R}_i^1(t,t')\Big(-\mathrm{i}\hat{x}_i(t)x_i(t')-R_i(t',t)\Big)\end{aligned} \tag{C.1.2}$$

From (2.3.32), we see that only equal time quadratic fields appear at first order, thus we write $\hat{C}_i^1(t,t') = \Delta^{-1}\hat{c}_i^1(t)\delta_{tt'}$ and $\hat{R}_i^1(t,t') = \Delta^{-1}\hat{r}_i^1(t)\delta_{tt'}$. To simplify, we switch to $\delta x_i(t)$, $\delta\hat{x}_i(t)$ and $\delta\phi_i = \phi_i - \langle\phi_i\rangle$ and we also introduce $\delta C_i(t,t') = C_i(t,t') - \mu_i(t)\mu_i(t')$ and $\delta R_i(t',t) = R_i(t',t) + \mathrm{i}\hat{\mu}_i(t)\mu_i(t')$. We subtract the mean as in (2.3.42) to obtain

$$\begin{aligned}\delta\,\frac{d\Xi_\alpha}{d\alpha} = \;\Delta\sum_{it}\Big[&\delta x_i(t)a_i(t) - \mathrm{i}\delta\hat{x}_i(t)b_i(t) - \mathrm{i}\hat{\mu}_i(t)\delta\phi_i(t) - \mathrm{i}\big(\delta\hat{x}_i(t)\delta\phi_i(t) - \mathrm{i}\delta R_i(t,t)\langle\partial_i\phi_i(t)\rangle\big) \\ &+\hat{c}_i^1(t)\big(\delta x_i(t)^2 - \delta C_i(t,t)\big) + \hat{r}_i^1(t)\big(-\mathrm{i}\delta\hat{x}_i(t)\delta x_i(t) - \delta R_i(t,t)\big)\Big]\end{aligned} \tag{C.1.3}$$

where we have introduced $b_i(t) = \langle\phi_i(t)\rangle + l_i^1(t) + \mu_i(t)\hat{r}_i^1(t)$ and $a_i(t) = \psi_i^1(t) + 2\mu_i(t)\hat{c}_i^1(t) - \mathrm{i}\hat{\mu}_i(t)\hat{r}_i(t)$; note that $b_i(t) \equiv 0$ as can be verified from (2.3.32b) and (2.3.32d).

The second order correction to the Plefka free energy is $G^2 = \langle(\delta\partial_\alpha\Xi_\alpha)^2\rangle/2$: the square will give

correlations of terms at different sites i and j and different times t and t'. Let us write G^2 explicitly

$$\begin{aligned}
G^2 \;=\; & \frac{\Delta^2}{2}\sum_{ijtt'}\Big[a_i(t)a_j(t')\langle\delta x_i(t)\delta x_j(t')\rangle - \hat{\mu}_i(t)\hat{\mu}_j(t')\langle\delta\phi_i(t)\delta\phi_j(t')\rangle - \langle\delta\hat{x}_i(t)\delta\hat{x}_j(t')\delta\phi_i(t)\delta\phi_j(t')\rangle \\
& -\delta R_i(t,t)\langle\partial_i\phi_i(t)\rangle\delta R_j(t',t')\langle\partial_j\phi_j(t)\rangle + \hat{c}_i(t)\hat{c}_j(t')\langle(\delta x_i^2(t)-\delta C_i(t,t))(\delta x_j(t')^2-\delta C_j(t',t'))\rangle \\
& +\hat{r}_i(t)\hat{r}_j(t')\langle(-\mathrm{i}\delta\hat{x}_i(t)\delta x_i(t)-\delta R_i(t,t))(-\mathrm{i}\delta\hat{x}_j(t')\delta x_j(t')-\delta R_j(t',t'))\rangle\Big] \\
& -\Delta^2\sum_{ijtt'}\Big[a_i(t)\mathrm{i}\hat{\mu}_j(t')\langle\delta x_i(t)\delta\phi_j(t')\rangle + a_i(t)\mathrm{i}\langle\delta x_i(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle + \hat{\mu}_j(t')\langle\delta\hat{x}_i(t)\delta\phi_i(t)\delta\phi_j(t')\rangle \\
& +\hat{c}_i(t)\mathrm{i}\hat{\mu}_j(t')\langle\delta x_i^2(t)\delta\phi_j(t')\rangle + \hat{r}_i(t)\mathrm{i}\hat{\mu}_j(t')\langle-\mathrm{i}\delta\hat{x}_i(t)\delta x_i(t)\delta\phi_j(t')\rangle + \mathrm{i}\hat{c}_i(t)(\langle\delta x_i^2(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle \\
& -\mathrm{i}\delta R_j(t',t')\langle\partial_j\phi_j(t')\rangle\delta C_i(t,t)) + \mathrm{i}\hat{r}_i(t)(\langle-\mathrm{i}\delta x_i(t)\delta\hat{x}_i(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle - \delta R_i(t,t)\langle\delta\hat{x}_j(t')\delta\phi_j(t')\rangle) \\
& -\hat{c}_i(t)\hat{r}_j(t')\langle(\delta x_i^2(t)-\delta C_i(t,t))(-\mathrm{i}\delta\hat{x}_j(t')\delta x_j(t')-\delta R_j(t',t'))\rangle\Big] \qquad \text{(C.1.4)}
\end{aligned}$$

where the first 3 lines refer to the quadratic terms while the last 4 lines give the cross-terms. The averages can be further developed by Wick's theorem (as at $\alpha = 0$ the statistics is Gaussian), thus one has

$$\langle\delta x_i(t)\delta x_j(t')\rangle = \delta_{ij}C_i(t,t') \qquad \text{(C.1.5a)}$$

$$\langle\delta\phi_i(t)\delta\phi_j(t')\rangle = \langle\phi_i(t)\phi_j(t')\rangle - \langle\phi_i(t)\rangle\langle\phi_j(t')\rangle \qquad \text{(C.1.5b)}$$

$$\begin{aligned}
\langle\delta\hat{x}_i(t)\delta\hat{x}_j(t')\delta\phi_i(t)\delta\phi_j(t')\rangle = & -\delta_{ij}\delta B_i(t,t')\langle\delta\phi_i(t)\delta\phi_i(t')\rangle + \mathrm{i}\delta R_i(t,t)\mathrm{i}\delta R_j(t,t')\langle(\partial_i\partial_j\phi_i(t))\delta\phi_j(t')\rangle \\
& + \mathrm{i}\delta R_i(t,t)\mathrm{i}\delta R_j(t',t')\langle\partial_i\phi_i(t)\partial_j\phi_j(t')\rangle + \mathrm{i}\delta R_i(t',t)\mathrm{i}\delta R_j(t,t')\langle\partial_j\phi_i(t)\partial_i\phi_j(t')\rangle \\
& + \mathrm{i}\delta R_i(t',t)\mathrm{i}\delta R_j(t',t')\langle\delta\phi_i(t)\partial_i\partial_j\phi_j(t')\rangle \qquad \text{(C.1.5c)}
\end{aligned}$$

$$\langle(\delta x_i^2(t)-\delta C_i(t,t))(\delta x_j(t')^2-\delta C_j(t',t'))\rangle = 2\delta_{ij}\delta C_i(t,t')^2 \qquad \text{(C.1.5d)}$$

$$\langle(-\mathrm{i}\delta\hat{x}_i(t)\delta x_i(t)-\delta R_i(t,t))(-\mathrm{i}\delta\hat{x}_j(t')\delta x_j(t')-\delta R_j(t',t'))\rangle = \delta_{ij}[\delta B_i(t,t')\delta C_i(t,t') + \delta R_i(t,t')\delta R_i(t',t)] \qquad \text{(C.1.5e)}$$

$$\langle\delta x_i(t)\delta\phi_j(t')\rangle = C_i(t,t')\langle\partial_i\phi_j(t')\rangle \qquad \text{(C.1.5f)}$$

$$\langle\delta x_i(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle = \delta C_i(t,t')\mathrm{i}\delta R_j(t',t')\langle\partial_i\partial_j\phi_j(t')\rangle \qquad \text{(C.1.5g)}$$

$$\langle\delta\hat{x}_i(t)\delta\phi_i(t)\delta\phi_j(t')\rangle = \mathrm{i}\delta R_i(t,t)\langle\partial_i\phi_i(t)\delta\phi_j(t')\rangle + \mathrm{i}\delta R_i(t',t)\langle\delta\phi_i(t)\partial_i\phi_j(t')\rangle \qquad \text{(C.1.5h)}$$

$$\langle\delta x_i^2(t)\delta\phi_j(t')\rangle = \delta C_i(t,t')^2\langle\partial_i^2\phi_j(t')\rangle \qquad \text{(C.1.5i)}$$

$$\langle-\mathrm{i}\delta\hat{x}_i(t)\delta x_i(t)\delta\phi_j(t')\rangle = \delta R_i(t',t)\delta C_i(t',t)\langle\partial_i^2\phi_j(t')\rangle \qquad \text{(C.1.5j)}$$

$$\begin{aligned}
\langle\delta x_i^2(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle = & \delta C_i(t,t)\mathrm{i}\delta R_j(t',t')\langle\partial_j\phi_j(t')\rangle + 2\delta_{ij}\mathrm{i}\delta R_i(t,t')\delta C_i(t,t')\langle\partial_i\phi_i(t')\rangle \\
& + \delta C_i(t,t')^2\mathrm{i}\delta R_j(t',t')\langle\partial_i^2\partial_j\phi_j(t')\rangle \qquad \text{(C.1.5k)}
\end{aligned}$$

$$\begin{aligned}
\langle-\mathrm{i}\delta x_i(t)\delta\hat{x}_i(t)\delta\hat{x}_j(t')\delta\phi_j(t')\rangle = & \delta R_i(t,t)\mathrm{i}\delta R_j(t',t')\langle\partial_j\phi_j(t')\rangle + \delta_{ij}\delta R_i(t,t')\mathrm{i}\delta R_i(t',t)\langle\partial_i\phi_i(t')\rangle \\
& + \mathrm{i}\delta_{ij}\delta B_i(t,t')\delta C_i(t,t')\langle\partial_i\phi_i(t')\rangle + \delta C_i(t,t')\delta R_i(t',t)\mathrm{i}\delta R_j(t',t')\langle\partial_i^2\partial_j\phi_j(t')\rangle \qquad \text{(C.1.5l)}
\end{aligned}$$

$$\langle(\delta x_i^2(t)-\delta C_i(t,t))(-\mathrm{i}\delta\hat{x}_j(t')\delta x_j(t')-\delta R_j(t',t'))\rangle = 2\delta_{ij}\delta C_i(t,t')\delta R_i(t,t') \qquad \text{(C.1.5m)}$$

where we have used (2.3.28) and $\langle \delta\phi_i(t)\rangle = 0$.

One now needs the derivatives of (C.1.4) w.r.t. $\mu_i(t)$, $\hat{\mu}_i(t)$, $C_i(t,t')$, $R_i(t,t')$ and $B_i(t,t')$ as stated in (2.3.43). The evaluation is simpler if we use that in the final solution we will only require these derivatives at the causal solution, where $\hat{\mu}_i(t) = 0$ and $\delta R_i(t,t') = 0$ for $t' \geq t$ - these imply also $a_i(t) = 0$ and $\hat{c}_i(t) = 0$.

$$\psi_i^2(t) = 0 \tag{C.1.6}$$

$$\begin{aligned} l_i^2(t) = {} & 2\Delta \sum_{jt'} \Big[\delta R_j(t,t') \left(-\langle \delta\phi_j(t')\partial_j\phi_i(t)\rangle + \langle \partial_j^2\phi_i(t)\rangle \delta C_j(t,t')\langle \partial_j\phi_j(t')\rangle \right) \\ & + \Delta\mu_i(t)\delta R_j(t,t')\langle (\partial_i\partial_j\phi_i(t))\delta\phi_j(t')\rangle + \Delta\mu_i(t')\delta R_j(t,t')\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle \\ & - \Delta\langle \partial_i\phi_i(t')\rangle\langle \partial_i\phi_i(t)\rangle\mu_i(t')\delta R_i(t,t') - \Delta\langle \partial_j\phi_j\rangle\delta C_j(t,t')\mu_i(t)\delta R_j(t,t')\langle \partial_j^2\partial_i\phi_i(t)\rangle \Big] \end{aligned} \tag{C.1.7}$$

$$\hat{C}_i^2(t,t') = 0 \tag{C.1.8}$$

$$\begin{aligned} \hat{R}_i^2(t,t') = {} & 2\sum_j \Big[-\delta R_j(t,t')\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle + \langle \partial_i\phi_i(t')\rangle\langle \partial_i\phi_i(t)\rangle\delta R_i(t,t') + \\ & -\delta R_j(t,t')\langle \partial_i\partial_j\phi_i(t)\delta\phi_j(t')\rangle + \langle \partial_j\phi_j(t')\rangle\delta C_j(t,t')\delta R_j(t,t')\langle \partial_j^2\partial_i\phi_i(t)\rangle \Big] \end{aligned} \tag{C.1.9}$$

$$\hat{B}_i^2(t,t') = -\langle \delta\phi_i(t)\delta\phi_i(t')\rangle + \langle \partial_i\phi_i(t)\rangle\langle \partial_i\phi_i(t')\rangle\delta C_i(t,t') \tag{C.1.10}$$

These fields, multiplied by $-\alpha^2$, give the second order contributions to the effective non-interacting dynamical action. Then the dynamical equation up to α^2 order can be written, from (2.2.14) and (2.3.17), in the form

$$\begin{aligned} \frac{dx_i(t)}{dt} = {} & -\lambda_i x_i(t) + \alpha\langle \phi_i(t)\rangle + \alpha\langle \partial_i\phi_i(t)\rangle\delta x_i(t) \\ & + \alpha^2 \Big[\int_0^t dt' \sum_j R_j(t,t')\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle\delta x_i(t') \\ & - \int_0^t dt' R_i(t,t')\langle \partial_i\phi_i(t)\rangle\langle \partial_i\phi_i(t')\rangle\delta x_i(t') \\ & + \int_0^t dt' \sum_j R_j(t,t')(\langle \partial_j\phi_i(t)\delta\phi_j(t')\rangle - \langle \partial_j^2\phi_i(t)\rangle\delta C_j(t,t')\langle \partial_j\phi_j(t')\rangle) \\ & + \int_0^t dt' \sum_j (\langle \partial_i\partial_j\phi_i(t)\delta\phi_j(t')\rangle - \delta C_j(t,t')\langle \partial_j\phi_j(t')\rangle\langle \partial_j^2\partial_i\phi_i(t)\rangle) \\ & R_j(t,t')\delta x_i(t) \Big] + \xi_i(t) + \chi_i(t) \end{aligned} \tag{C.1.11}$$

where the effective noise has correlator

$$\langle \chi_i(t)\chi_i(t')\rangle = \frac{\alpha^2}{2}(\langle \delta\phi_i(t)\delta\phi_i(t')\rangle - \langle \partial_i\phi_i(t)\rangle\langle \partial_i\phi_i(t')\rangle\delta C_i(t,t')) \tag{C.1.12}$$

For the sake of brevity we have dropped all $\boldsymbol{x}$-dependencies above, writing e.g. $\partial_i\phi_i(t) = \partial\phi_i(\boldsymbol{x}(t))/\partial x_i(t)$ and $\phi_i(t) = \phi_i(\boldsymbol{x}(t))$.

Compared to (2.3.48) in the main text, there are a number of additional terms. The last term in the first line is the linearization of the self-interaction already familiar from our generic first order result (2.3.36). This systematic effect of the self-interaction is correspondingly removed from the effective noise $\chi_i(t)$, whose correlator (C.1.12) is easily shown to be the correlation function of $\delta\tilde{\phi}_i \equiv \delta\phi_i - \langle\partial_i\phi_i\rangle\delta x_i$. This is the genuinely interacting part of the drift, i.e. the one that is not captured in the first line of (C.1.11). The third line similarly subtracts off the self-interaction term from the main memory term in the second line.

The fourth line of (C.1.11) is a contribution that is independent of the specific history of x_i; instead it involves a time integral of *averages* over fluctuation statistics in the past. It can again be written in terms of $\delta\tilde{\phi}_i$, with the coefficients in brackets after $R_j(t,t')$ equal to $\langle\partial_j\phi_i(t)\delta\tilde{\phi}_j(t')\rangle$. The coefficient in front of $\delta x_i(t)$ in the fifth and sixth line of (C.1.11) has an analogous form, as $\langle\partial_i\partial_j\phi_i(t)\delta\tilde{\phi}_j(t')\rangle$. The overall contribution from the fourth, fifth and sixth line of (C.1.11) can be cast as

$$\alpha^2\int_0^t dt'\sum_j R_j(t,t')(\langle\partial_j\phi_i(t)\delta\tilde{\phi}_j(t')\rangle + \delta x_i(t)\langle\partial_i\partial_j\phi_i(t)\delta\tilde{\phi}_j(t')\rangle) \tag{C.1.13}$$

This has a fairly straightforward interpretation: a fluctuation in the drift of variable j ($\delta\tilde{\phi}_j(t')$) that changes $x_j(t')$ is propagated forward to time t by $R_j(t,t')$ and then affects the drift $\tilde{\phi}_i(t)$ including the linearized dependence on x_i.

It is interesting to note that all of the additional terms disappear if there are no self-interactions ($\partial_i\phi_i = 0$), except for the first term in the fourth line of (C.1.11). The latter vanishes if one makes in addition the assumption that interactions are additive in the variables, as then $\partial_j\phi_i$ depends only on x_j and so is independent of $\delta\tilde{\phi}_j$ if there are no self-interactions. In the generic case of non-additive interactions, this term remains. In particular, it gives a correction to the time evolution of the means

$$\frac{d\mu_i(t)}{dt} = -\lambda_i\mu_i(t) + \alpha\langle\phi_i(t)\rangle + \alpha^2\int_0^t dt'\sum_j R_j(t,t')\langle\partial_j\phi_i(t)\delta\tilde{\phi}_j(t')\rangle \tag{C.1.14}$$

In an exact theory, only the first two terms are present, so that the last one has to be interpreted as correcting for the fact that the Plefka expansion produces an approximating distribution where all variables are decoupled. For the case of additive interactions, no such correction appears because $\langle\phi_i\rangle$ is then a sum of averages over single variables.

C.1.1 *p*-spin model

For the special case of the p-spin spherical model, we have verified that the above equations reproduce those derived by other means by Biroli [52]. The correction to the mean dynamics

vanishes for $p = 2$ as expected, as the interactions are then additive, but is nonzero for $p > 2$ where the drift involves products of variables.

The detailed calculation is as follows. Our single site dynamics without self-interactions (other than λ_i) but non-additive interactions, at $\alpha = 1$, is given by

$$\begin{aligned}\frac{dx_i(t)}{dt} &= -\lambda_i x_i(t) + \langle \phi_i(t)\rangle + \xi_i(t) + \chi_i(t) \\ &+ \int_0^t dt' \sum_j R_j(t,t')(\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle \delta x_i(t') + \langle \partial_j\phi_i(t)\delta\phi_j(t')\rangle)\end{aligned} \tag{C.1.15}$$

with

$$\hat{B}_i^{\text{eff}}(t,t') = \langle \delta\phi_i(t)\delta\phi_i(t')\rangle \tag{C.1.16}$$

We set $\lambda_i = \lambda$ for any i. To compare (C.1.15) with the analysis in [52], we set

$$\phi_i(t) = -\frac{\partial H}{\partial x_i(t)} = \sum_{1\le i'_1<\ldots<i'_{p-1}\le N} J_{i'_1,\ldots,i,\ldots,i'_{p-1}} x_{i'_1}(t)\ldots x_{i'_{p-1}}(t) \qquad \{i'\} \neq i \tag{C.1.17}$$

as in a p-spin model, where $\boldsymbol{x}$s play the role of spins. Given this definition, we calculate the derivatives and averages appearing in (C.1.15). We have

$$\langle \phi_i(t)\rangle = \sum_{1\le i'_1<\ldots<i'_{p-1}\le N} J_{i'_1,\ldots,i,\ldots,i'_{p-1}} \mu_{i'_1}(t)\ldots\mu_{i'_{p-1}}(t) \tag{C.1.18}$$

(note that correlations are only diagonal) and

$$\partial_j\phi_i(t) = \sum_{\{i'\}} J_{i'_1,\ldots,i,j,\ldots,i'_{p-2}} x_{i'_1}(t)\ldots x_{i'_{p-2}}(t) \tag{C.1.19}$$

For brevity we have set $\{i'\} = 1 \le i'_1 < \ldots < i'_{p-2} \le N$, with $\{i'\} \neq i, j$. Similarly

$$\partial_i\phi_j(t') = \sum_{\{j'\}} J_{j'_1,\ldots,i,j\ldots,j'_{p-2}} x_{j'_1}(t')\ldots x_{j'_{p-2}}(t') \tag{C.1.20}$$

with $\{j'\} = 1 \le j'_1 < \ldots < j'_{p-2} \le N$, with $\{j'\} \neq i, j$. The aim is to show that the expressions for the dynamics of means, correlations and responses one can derive from (C.1.15) are consistent with Biroli's results (see equations 44-46 in [52]).

We want to start with $\hat{R}^{\text{eff}}(t,t') = \sum_j R_j(t,t')\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle$, which can be expressed as follows

$$\begin{aligned}&\sum_j R_j(t,t')\langle \partial_j\phi_i(t)\partial_i\phi_j(t')\rangle = \\ &= \sum_j\sum_{\{i'\}}\sum_{\{j'\}} R_j(t,t')\langle J_{i'_1,\ldots,i,j,\ldots,i'_{p-2}} J_{j'_1,\ldots,i,j,\ldots,j'_{p-2}} x_{i'_1}(t)\ldots x_{i'_{p-2}}(t) x_{j'_1}(t')\ldots x_{j'_{p-2}}(t')\rangle = \\ &= \sum_j\sum_{\{i'\}} R_j(t,t')\langle J^2_{i'_1,\ldots,i,j,\ldots,i'_{p-2}}\rangle\langle x_{i'_1}(t)\ldots x_{i'_{p-2}}(t) x_{i'_1}(t')\ldots x_{i'_{p-2}}(t')\rangle = \\ &= \sum_j\sum_{\{i'\}} R_j(t,t')\frac{p!}{2N^{p-1}} C_{i'_1}(t,t')\ldots C_{i'_{p-2}}(t,t') = R(t,t')\frac{p!}{2}\frac{1}{(p-2)!}C(t,t')^{p-2}\end{aligned} \tag{C.1.21}$$

thus

$$\hat{R}^{\text{eff}}(t,t') = \frac{p(p-1)}{2} R(t,t') C(t,t')^{p-2} \tag{C.1.22}$$

where again we have used the fact that only the diagonal elements of correlations are non-zero, thus $\{i'\} = \{j'\}$. The $1/(p-2)!$ is needed as when one defines C^{p-2} the order of indices is not fixed as we have in all these sums. Also, in developing the averages one considers the thermodynamic limit behaviour of weak, long-range couplings, i.e. $\langle J^2_{i'_1,..,i,j,..,i'_{p-2}} \rangle = p!/2N^{p-1}$. Underlying there is the self-averaging argument, similarly to the linear model in the main text (see (2.4.4) and (2.4.5)), yet we use the quenched average as a shortcut to get the result straightforwardly. Responses and correlations are expected to be self-averaging while μ_i can be dependent on the site, as it is in the spin glass phase; the average $C(t,t')$ and $R(t,t')$ are defined as in section 2.4.2. Given $\hat{R}^{\text{eff}}(t,t')$ in the thermodynamic limit, we can get from (C.1.15) a differential equation for the average response (similarly to the derivation of (2.3.19) in the main text)

$$\partial_t R(t,t') = -\lambda R(t,t') + \delta(t-t') + \frac{p(p-1)}{2} \int_{t'}^{t} dt'' R(t,t'') C(t,t'')^{p-2} R(t'',t) \tag{C.1.23}$$

exactly expression 45 in [52] (where $\mu = p\beta^2/2$, $\beta = 1$). For the temporal evolution of the correlations we need also $\hat{B}^{\text{eff}}_i(t,t') = \langle \delta\phi_i(t)\delta\phi_i(t') \rangle$, thus

$$\langle \delta\phi_i(t)\delta\phi_i(t') \rangle = \sum_{\{i'\}} \sum_{\{i''\}} \langle J_{i'_1,..,i,..,i'_{p-1}} J_{i''_1,..,i,..,i''_{p-1}} \delta(x_{i'_1}(t)...x_{i'_{p-1}}(t)) \delta(x_{i''_1}(t')...x_{i''_{p-1}}(t')) \rangle \tag{C.1.24}$$

with $\{i'\} = 1 \leq i'_1 < ... < i'_{p-1} \leq N$, $\{i'\} \neq i$ and similarly $\{i''\} = 1 \leq i''_1 < ... < i''_{p-1} \leq N$, $\{i''\} \neq i$. Only terms for which each i' is equal to the corresponding i'' in the right ordering actually give a non-zero contribution as connected correlations have only diagonal terms (otherwise $\delta(x_{i'_1}(t)...x_{i'_{p-1}}(t))$ and $\delta(x_{i''_1}(t')...x_{i''_{p-1}}(t')$ factorize and their average vanishes).
Then one has

$$\begin{aligned}
\langle \delta\phi_i(t)\delta\phi_i(t') \rangle &= \sum_{\{i'\}} \langle J^2_{i'_1,..,i,..,i'_{p-1}} \delta(x_{i'_1}(t)...x_{i'_{p-1}}(t)) \delta(x_{i'_1}(t')...x_{i'_{p-1}}(t')) \rangle = \\
&= \sum_{\{i'\}} \langle J^2_{i'_1,..,i,..,i'_{p-1}} \rangle \Big(\langle x_{i'_1}(t)...x_{i'_{p-1}}(t) x_{i'_1}(t')...x_{i'_{p-1}}(t') \rangle - \mu_{i'_1}(t)...\mu_{i'_{p-1}}(t)\mu_{i'_1}(t')...\mu_{i'_{p-1}}(t') \Big) \\
&= \frac{p!}{2N^{p-1}} \sum_{\{i'\}} \Big(C_{i'_1}(t,t')...C_{i'_{p-1}}(t,t') - \mu_{i'_1}(t)...\mu_{i'_{p-1}}(t)\mu_{i'_1}(t')...\mu_{i'_{p-1}}(t') \Big) = \\
&= \frac{p!}{2} \frac{1}{(p-1)!} \Big(C(t,t')^{p-1} - Q(t,t')^{p-1} \Big)
\end{aligned} \tag{C.1.25}$$

where we have introduced $Q(t,t') = \frac{1}{N} \sum_{i=1}^{N} \mu_i(t)\mu_i(t')$. As a result

$$\hat{B}^{\text{eff}}_i(t,t') = \frac{p}{2} (C(t,t')^{p-1} - Q(t,t')^{p-1}) \tag{C.1.26}$$

Thus, noting that the connected correlation $\delta C(t,t') = C(t,t') - Q(t,t')$, the final dynamical equation one can derive from (C.1.15) (in analogy with (2.3.20)) is given by

$$\begin{aligned}\partial_t(C(t,t') - Q(t,t')) = -\lambda(C(t,t') - Q(t,t')) + \tfrac{p(p-1)}{2}\int_{t'}^{t} dt'' R(t,t'')C(t,t'')^{p-2}(C(t'',t') - Q(t'',t')) \\ + \tfrac{p}{2}\int_0^{t'} dt'' R(t',t'')(C(t,t'')^{p-1} - Q(t,t'')^{p-1}) + \Sigma R(t',t)\end{aligned} \tag{C.1.27}$$

exactly equation 44 in [52], where $\Sigma = 2$.

For the means, one has the additional contribution

$$\begin{aligned}&\sum_j R_j(t,t')\langle \partial_j \phi_i(t)\delta\phi_j(t')\rangle = \\ =\ &\sum_j \sum_{\{i'\}}\sum_{\{j'\}} R_j(t,t')\langle J_{i'_1,...,i,j..,i'_{p-2}} J_{j'_1,...,j,..,j'_{p-1}} x_{i'_1}(t)...x_{i'_{p-2}}(t)\delta(x_{j'_1}(t')...x_{j'_{p-1}}(t'))\rangle = \\ =\ &\sum_j \sum_{\{i'\}} R_j(t,t')\langle J^2_{i'_1,...,i,j..,i'_{p-2}}\rangle\langle x_{i'_1}(t)...x_{i'_{p-2}}(t)x_{i'_1}(t')...x_{i'_{p-2}}(t')x_i(t')\rangle \\ -\ &\mu_{i'_1}(t)...\mu_{i'_{p-2}}(t)\mu_{i'_1}(t')...\mu_{i'_{p-2}}(t')\mu_i(t') = \\ =\ &\sum_j \sum_{\{i'\}} R_j(t,t')\frac{p!}{2N^{p-1}}\Big(C_{i'_1}(t,t')...C_{i'_{p-2}}(t,t') - \mu_{i'_1}(t)...\mu_{i'_{p-2}}(t)\mu_{i'_1}(t')...\mu_{i'_{p-2}}(t')\Big)\mu_i(t') = \\ =\ &R(t,t')\frac{p!}{2(p-2)!}\Big(C(t,t')^{p-2} - Q(t,t')^{p-2}\Big)\mu_i(t') = \\ =\ &\frac{p(p-1)}{2}R(t,t')\Big(C(t,t')^{p-2} - Q(t,t')^{p-2}\Big)\mu_i(t')\end{aligned} \tag{C.1.28}$$

with $\{i'\} = 1 \le i'_1 < ... < i'_{p-2} \le N$, $\{i'\} \neq i, j$, and $\{j'\} = 1 \le j'_1 < ... < j'_{p-1} \le N$, $\{j'\} \neq j$ (note that there is one $j' = i$; the only terms to survive are the ones for which the $p-1$ indices $\{j'\} = \{i'\} + i$). As a result, the expression for the mean is

$$\begin{aligned}\partial_t \mu_i(t) \quad &= -\lambda\mu_i(t) + \textstyle\sum_{1\le i_1<...<i_{p-1}\le N} J_{i_1,...,i,..,i_{p-1}}\mu_{i_1}(t)...\mu_{i_{p-1}}(t) \\ &+ \tfrac{p(p-1)}{2}\int_0^t dt' R(t,t')\Big(C(t,t')^{p-2} - Q(t,t')^{p-2}\Big)\mu_i(t')\end{aligned} \tag{C.1.29}$$

which is the same as expression 46 in [52].

We finally remark the match also with the effective disorder-averaged Langevin equation derived in [53]

$$\partial_t x(t) = -\lambda x(t) + \frac{1}{2}p(p-1)\int_0^t dt' R(t,t')C(t,t')^{p-2}x(t') + \xi(t) + \chi(t) \tag{C.1.30}$$

where $\mu(t) = 0$ and noise correlations are given by

$$\langle \xi(t)\xi(t')\rangle = \Sigma\delta(t-t') \tag{C.1.31}$$

and

$$\langle \chi(t)\chi(t')\rangle = \frac{p}{2}C(t,t')^{p-1} \tag{C.1.32}$$

Expression (C.1.30) applies in the paramagnetic phase and there (as all $\mu_i = 0$) the extended Plefka effective dynamics (2.3.17) is consistent with this result in the thermodynamic limit, i.e. when one can drop the i index as all sites are equivalent and consider average correlations and responses. The correspondence can be verified by taking the effective fields for the $p-$spin model with vanishing mean $\mu_i(t) = 0$ (so that also $Q(t, t') = 0$), i.e. $l_i^{\text{eff}}(t) = 0$ and expressions (C.1.22) for $\hat{R}^{\text{eff}}(t, t')$ and (C.1.26) for $\hat{B}_i^{\text{eff}}(t, t')$.

CHAPTER 3

Inference for dynamics: the Extended Plefka Expansion with hidden nodes

The related paper has been submitted to Journal of Statistical Mechanics: Theory and Experiment and currently under review [56].

> Most local epistemologies - personal and cultural - continually err, alas, in confusing map with territory and in assuming that the rules for drawing maps are immanent in the nature of that which is being represented in the map. [...] The map is never the territory, but it is sometimes useful to discuss how map differs from hypothetical territory.
>
> *Gregory Bateson, "Mind and Nature"*

3.1 Introduction

The problem of reconstructing the time evolution of a system given some measurements of its dynamics has seen much recent interest in the statistical physics community [57–59]. Given a temporal sequence of observed variables, the task is to infer the states of other variables that are not observed, based on knowledge of the interaction parameters. Techniques for tackling this problem could have a significant impact in systems biology, where dealing with missing variables and experimental limitations requires the development of novel inference frameworks, to enable quantitative modelling on the basis of experimental data [60–62].

In machine learning and pattern recognition, the problem of inference from data has been addressed using e.g. state space models [16] that introduce "hidden" variables playing the role of

unobserved states. The best-known example of this type is probably the Kalman filter [63], for the case of stochastic linear dynamics of continuous degrees of freedom (d.o.f.). We study such a dynamics here, with interaction parameters drawn at random from a Gaussian distribution. A wide range of microscopic biological processes can be captured qualitatively even by this simplified setting, such as inter- and intra-cellular biochemical networks where interactions parameters are given by reaction rates and one may be interested in analyzing fluctuations around a steady state [64]. The joint distribution over observed and hidden states is Gaussian in our model and this allows us to analyze the posterior statistics of the hidden dynamics. The second order statistics in particular tells us how the observed dynamics constrains the unknown hidden dynamics, with the posterior variances quantifying the degree of uncertainty in hidden state prediction. As our scenario is analytically tractable, we can study the dependence of this prediction error on key parameters and thus shed light on the accuracy of the inference process, similarly to what has been done for learning in a linear perceptron [65, 66].

We tackle the inference problem for our setting by means of the Extended Plefka Expansion that we have developed in chapter 2, a dynamical mean field theory for continuous degrees of freedom. This should become exact in the thermodynamic limit of a large network. It is also general enough to allow us to treat a wide range of dynamical interactions, i.e. couplings of any symmetry, so that we can probe both equilibrium and non-equilibrium networks with only partially observed dynamics. This flexibility makes our approach in principle widely applicable to real data from systems driven out of equilibrium by fluxes, with biological networks being a case in point as illustrated e.g. by studies on non-equilibrium steady states for reaction fluxes [67] and “near" symmetry features in metabolic networks [68].

Inference problems for non-equilibrium systems have been already investigated for neural data, using either two-state units (Ising spins) [59] or deterministic continuous-valued hidden units [69]. These studies were motivated by modelling populations of neurons, and concentrated on finding learning rules, while here we study continuous d.o.f. with random linear interactions in the large network limit and our focus is on understanding the prediction accuracy for hidden states theoretically and from a macroscopic point of view, including its dependence on key system parameters. State inference is indeed a complementary step of parameter estimation towards the overall prediction of an unknown dynamics. For example existing algorithmic tools, such as Expectation Propagation [54, 55], iterate between estimating the hidden states (given the couplings) and estimating the couplings (given the updated states).

This chapter is organized as follows. In section 3.2 we set our model in terms of a set of dynamical

equations for observed and hidden continuous variables interacting via linear Gaussian couplings. A path integral representation of the likelihood (see e.g. [59]) provides the starting point for the Extended Plefka Expansion. The latter produces an effectively non-interacting approximation of the original dynamics, more specifically a Gaussian posterior probability that is factorized over hidden nodes but incorporates the hidden-to-observed couplings. In section 3.2.1 we derive equations for the posterior means, which give the best estimate of the hidden dynamics. In section 3.2.2 we focus then on posterior second moments, i.e. hidden-to-hidden correlations, hidden responses and auxiliary correlations, in the stationary regime. In section 3.2.3 we consider the thermodynamic limit of infinite network size, shifting from local correlations to their macroscopic average across the network. The exactness, at least in some limit cases, of the extended Plefka predictions is shown in section 3.2.4 by comparison with expressions we will systematically derive in chapter 4 by Kalman filter and Random Matrix Theory (RMT) methods. In section 3.3.1 we determine the relevant dimensionless parameters governing prediction accuracy, namely the ratio between the numbers of observed and hidden nodes, the degree of symmetry of the hidden interactions, and the amplitudes of the hidden-to-hidden and hidden-to-observed interactions relative to the decay constant of the internal hidden dynamics. We identify in this parameter space critical points for the inference error and study its behaviour in the critical regions by appealing to a scaling analysis (sections 3.3.2, 3.3.3, 3.3.4). We also consider the temporal correlations in the posterior dynamics of the hidden nodes, revealing interesting long-time behaviour in the critical regions and we study the posterior relaxation times for general parameter settings in section 3.4.

3.2 Extended Plefka Expansion with hidden nodes

We consider a generic network where only the dynamics of a subnetwork of nodes is observed while the others are hidden and form what we call the "bulk". Such a situation could arise because the subnetwork nodes are more precisely characterized from the theoretical point of view, or experimentally more accessible. We assume that we have noise-free data for the trajectories of the observed nodes. In biological contexts this is clearly a simplification as data is often available only at discrete time points or corrupted by noise. The constraint that only a part of the network can be observed is generic, on the other hand. In protein interaction networks, for example, only a few molecular species can typically be tagged biochemically in such a way that their concentrations can be tracked with reasonable accuracy [6, 70].

We use the indices $i, j = 1, ..., N^{\mathrm{b}}$ for the hidden or bulk variables and $a, b = 1, ..., N^{\mathrm{s}}$ for the observed or subnetwork nodes of the network, and generally use the superscripts s and b to distin-

guish variables relating to the observed and hidden sectors, respectively. Assuming that the hidden and observed variables x_i, x_a interact via linear couplings $\{J_{ij}\}$, $\{K_{ia}\}$, their dynamical evolution is described by

$$\partial_t x_i(t) = -\lambda x_i(t) + \sum_j J_{ij} x_j(t) + \sum_a K_{ia} x_a(t) + \xi_i(t) \tag{3.2.1a}$$

$$\partial_t x_a(t) = -\lambda x_a(t) + \sum_b J_{ab} x_b(t) + \sum_j K_{aj} x_j(t) + \xi_a(t) \tag{3.2.1b}$$

J_{ij} (respectively J_{ab}) denotes hidden-to-hidden (respectively observed-to-observed) interactions, while the coupling between observed and hidden variables is contained in K_{ia} and K_{aj}. We have included a self-interaction term with coefficient λ, which acts as a decay constant and provides the basic timescale of the dynamics. In more compact notation we can write

$$\partial_t x_i(t) = -\lambda x_i(t) + \phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \xi_i(t) \tag{3.2.2a}$$

$$\partial_t x_a(t) = -\lambda x_a(t) + \phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \xi_a(t) \tag{3.2.2b}$$

where $\phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) = \sum_j J_{ij} x_j(t) + \sum_a K_{ia} x_a(t)$ and $\phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) = \sum_b J_{ab} x_b(t) + \sum_j K_{aj} x_j(t)$ and $\boldsymbol{x}^{\mathrm{b}}(t)$ ($\boldsymbol{x}^{\mathrm{s}}(t)$) denotes the whole set of hidden (observed) values. The main assumption of this setup is that the trajectory described by (3.2.2b) is known in a finite time window $\{0, T\}$.

The dynamical noises ξ_i, ξ_a are Gaussian white noises with zero mean and diagonal covariances Σ_i, Σ_a

$$\langle \xi_i(t)\xi_j(t')\rangle = \Sigma_i \delta_{ij}\delta(t-t') \qquad \langle \xi_a(t)\xi_b(t')\rangle = \delta_{ab}\Sigma_a\delta(t-t') \tag{3.2.3}$$

where the diagonal structure is assumed in analogy with chapter 2.

After discretizing time in elementary time steps Δ, we can write the likelihood – in our case, the probability of a trajectory of the observed variables – using the Martin–Siggia–Rose–Janssen–De Dominicis (MSRJD) path integrals formalism [19–21] as

$$\begin{aligned}
P(\boldsymbol{x}^{\mathrm{s}}) &= \Big\langle \int D\boldsymbol{x}^{\mathrm{b}} \prod_{at} \delta(x_a(t+\Delta) - x_a(t) - \Delta[-\lambda x_a(t) + \phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \xi_a(t)]) \\
&\quad \prod_{it} \delta(x_i(t+\Delta) - x_i(t) - \Delta[-\lambda x_i(t) + \phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \xi_i(t)])\Big\rangle_{\boldsymbol{\xi}^{\mathrm{s}}, \boldsymbol{\xi}^{\mathrm{b}}} = \\
&= \Big\langle \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{s}} \exp\Big\{ \sum_{at} \mathrm{i}\hat{x}_a(t)(x_a(t+\Delta) - x_a(t) - \Delta[-\lambda x_a(t) + \phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) \\
&+ \xi_a(t)]) + \sum_{it} \mathrm{i}\hat{x}_i(t)(x_i(t+\Delta) - x_i(t) - \Delta[-\lambda x_i(t) + \phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \xi_i(t)])\Big\}\Big\rangle_{\boldsymbol{\xi}^{\mathrm{s}}, \boldsymbol{\xi}^{\mathrm{b}}} = \\
&= \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{s}}\, \mathrm{e}^{\mathcal{H}}
\end{aligned} \tag{3.2.4}$$

where $D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}}$ and $D\hat{\boldsymbol{x}}^{\mathrm{s}}$ are shorthands for the integrations $\prod_{it} dx_i(t) d\hat{x}_i(t)/2\pi$ and $\prod_{at} d\hat{x}_a(t)/2\pi$

respectively and the action $\mathcal{H}$

$$\begin{aligned}\mathcal{H} &= \sum_{at} \mathrm{i}\hat{x}_a(t)(x_a(t+\Delta) - x_a(t) - \Delta[-\lambda x_a(t) + \phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))]) - \frac{\Delta}{2}\Sigma_a \hat{x}_a(t)\hat{x}_a(t) \\ &\quad \sum_{it} \mathrm{i}\hat{x}_i(t)(x_i(t+\Delta) - x_i(t) - \Delta[-\lambda x_i(t) + \phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))]) - \frac{\Delta}{2}\Sigma_i \hat{x}_i(t)\hat{x}_i(t) \qquad (3.2.5)\end{aligned}$$

The $\hat{\boldsymbol{x}}^{\mathrm{s}}(t), \hat{\boldsymbol{x}}^{\mathrm{b}}(t)$ are auxiliary variables one introduces to represent the δ function enforcing the dynamics (3.2.1) in terms of an exponential. In the final step of (3.2.4) we have applied a standard Gaussian identity to average over the Gaussian noises, i.e.

$$\langle \mathrm{e}^{\mathrm{i}\Delta \hat{\boldsymbol{x}}^{\cdot T}\boldsymbol{\xi}^{\cdot}} \rangle_{\boldsymbol{\xi}^{\cdot}} = \mathrm{e}^{-\Delta\, \hat{\boldsymbol{x}}^{\cdot T}\boldsymbol{\Sigma}^{\cdot}\hat{\boldsymbol{x}}^{\cdot}/2} \qquad \cdot = \mathrm{s}, \mathrm{b} \qquad (3.2.6)$$

where the superscript $\cdot = \mathrm{s}, \mathrm{b}$ refers to both the subnetwork (s) and the bulk (b). We have used the Itô convention [71] to discretize the noise and $\xi_i(t)$ above is to be read as the average of the continuous-time noise over the time interval $[t, t+\Delta]$, with covariance

$$\langle \xi_i(t)\xi_i(t') \rangle = \frac{1}{\Delta}\Sigma_i \delta_{tt'} \qquad \langle \xi_a(t)\xi_a(t') \rangle = \frac{1}{\Delta}\Sigma_a \delta_{tt'} \qquad (3.2.7)$$

Here $\delta_{tt'}/\Delta$ is the discrete-time analogue of $\delta(t-t')$. Note that (3.2.4) can be viewed as a partition function that normalizes the posterior distribution of hidden trajectories given the observed trajectory. Because of the conditioning on the observations it is not equal to unity, so that one has to be careful not to rely on consequences – such as the vanishing of all moments of $\hat{\boldsymbol{x}}^{\mathrm{b}}$ and $\hat{\boldsymbol{x}}^{\mathrm{b}}$ – that would otherwise follow from this.

The essence of our approach is to treat the interacting terms $\phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))$ and $\phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))$ within the Extended Plefka Expansion, thus generalizing the approach in chapter 2 to the presence of observations. The first step is to decide what averages to fix as order parameters for the expansion; we choose the first and second moments of the fluctuating quantities (i.e. the ones we integrate over) $\boldsymbol{x}^{\mathrm{b}}(t)$, $\hat{\boldsymbol{x}}^{\mathrm{b}}(t)$ and $\hat{\boldsymbol{x}}^{\mathrm{s}}(t)$. Because we assume that the trajectory of the observed variables has been observed, the $\boldsymbol{x}^{\mathrm{s}}(t)$ are known $\forall t$ and do not need to be estimated. By analogy with the notation in chapter 2 we introduce shorthands for the quantities to be averaged, $\hat{\boldsymbol{m}}$, and for the order parameters they define, $\boldsymbol{m} = \langle \hat{\boldsymbol{m}} \rangle$

$$\hat{\boldsymbol{m}}^{\mathrm{b}} = \{\boldsymbol{x}^{\mathrm{b}}, -\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}, \boldsymbol{x}^{\mathrm{b}}\boldsymbol{x}^{\mathrm{b}}, -\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}\boldsymbol{x}^{\mathrm{b}}, \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}\} \qquad (3.2.8a)$$

$$\hat{\boldsymbol{m}}^{\mathrm{s}} = \{-\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}, \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}\} \qquad (3.2.8b)$$

$$\boldsymbol{m}^{\mathrm{b}} = \{\boldsymbol{\mu}^{\mathrm{b}}, -\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}}, \boldsymbol{C}^{\mathrm{bb}}, \boldsymbol{R}^{\mathrm{bb}}, \boldsymbol{B}^{\mathrm{bb}}\} \qquad (3.2.8c)$$

$$\boldsymbol{m}^{\mathrm{s}} = \{-\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{s}}, \boldsymbol{B}^{\mathrm{ss}}\} \qquad (3.2.8d)$$

where the components of $\boldsymbol{m}^{\mathrm{b}}$ and $\boldsymbol{m}^{\mathrm{s}}$ are explicitly defined as

$$\mu_i(t) = \langle x_i(t) \rangle \tag{3.2.9a}$$

$$\hat{\mu}_i(t) = \langle \hat{x}_i(t) \rangle \tag{3.2.9b}$$

$$\hat{\mu}_a(t) = \langle \hat{x}_a(t) \rangle \tag{3.2.9c}$$

$$C_i(t,t') = \langle x_i(t) x_i(t') \rangle \tag{3.2.9d}$$

$$R_i(t',t) = -\mathrm{i}\langle \hat{x}_i(t) x_i(t') \rangle \tag{3.2.9e}$$

$$B_i(t,t') = -\langle \hat{x}_i(t) \hat{x}_i(t') \rangle \tag{3.2.9f}$$

$$B_a(t,t') = -\langle \hat{x}_a(t) \hat{x}_a(t') \rangle \tag{3.2.9g}$$

These averages are the posterior moments over physical and auxiliary dynamical trajectories. Note that we keep only the diagonal second moments: this is crucial in order to obtain an effective non-interacting description from the Plefka expansion. This implies that different sites, including the observed a and hidden ones i, are effectively decoupled at the level of correlations and responses, the contribution of the mutual coupling being embedded in appropriately defined single site fields. The starting point of the expansion (see chapter 2) is in fact to augment the original partition function $P(\boldsymbol{x}^{\mathrm{s}})$ with field terms – denoted in our context $\boldsymbol{h}^{\mathrm{b}}_{\alpha}$ and $\boldsymbol{h}^{\mathrm{s}}_{\alpha}$ – conjugate to the chosen observables, and then consider the Legendre transform of the corresponding free energy

$$G_\alpha(\boldsymbol{m}^{\mathrm{s}}, \boldsymbol{m}^{\mathrm{b}}, \boldsymbol{x}^{\mathrm{s}}) = \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{s}}\, \mathrm{e}^{\Xi_\alpha} \tag{3.2.10}$$

where

$$\Xi_\alpha = \mathcal{H}_\alpha + \boldsymbol{h}^{\mathrm{b}}_{\alpha}(\hat{\boldsymbol{m}}^{\mathrm{b}} - \boldsymbol{m}^{\mathrm{b}}) + \boldsymbol{h}^{\mathrm{s}}_{\alpha}(\hat{\boldsymbol{m}}^{\mathrm{s}} - \boldsymbol{m}^{\mathrm{s}}) \tag{3.2.11}$$

The key of the Plefka expansion is the introduction of the parameter α here, which scales the interacting parts of the Hamiltonian. For $\alpha = 0$ one then has a non-interacting system and the Plefka expansion is then a perturbation expansion of G_α around this point, where one sets $\alpha = 1$ at the end to recover the full Hamiltonian $\mathcal{H} \equiv \mathcal{H}_1$.

In our model, Ξ_α is given explicitly by

$$\begin{aligned}
\Xi_\alpha \quad &= \sum_{it} \mathrm{i}\hat{x}_i(t)\big(x_i(t+\Delta) - x_i(t) + \Delta\lambda_i x_i(t) - \alpha\Delta\phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))\big) + \\
&+ \sum_{at} \mathrm{i}\hat{x}_a(t)\big(x_a(t+\Delta) - x_a(t) + \Delta\lambda_a x_a(t) - \alpha\Delta\phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t))\big) + \\
&+ \Delta \sum_{it} \psi_{i\alpha}(t)\big(x_i(t) - \mu_i(t)\big) - \Delta \sum_{it} l_{i\alpha}(t)\big(\mathrm{i}\hat{x}_i(t) - \mathrm{i}\hat{\mu}_i(t)\big) - \Delta \sum_{at} l_{a\alpha}(t)\big(\mathrm{i}\hat{x}_a(t) - \mathrm{i}\hat{\mu}_a(t)\big) + \\
&+ \Delta^2 \sum_{itt'} \hat{C}_{i\alpha}(t,t')\big(x_i(t)x_i(t') - C_i(t,t')\big) + \Delta^2 \sum_{itt'} \hat{R}_{i\alpha}(t,t')\big(- \mathrm{i}\hat{x}_i(t)x_i(t') - R_i(t',t)\big) +
\end{aligned}$$

$$+\frac{\Delta^2}{2}\sum_{itt'}\hat{B}_{i\alpha}(t,t')(-\hat{x}_i(t)\hat{x}_i(t')-B_i(t,t'))+\frac{\Delta^2}{2}\sum_{att'}\hat{B}_{a\alpha}(t,t')(-\hat{x}_a(t)\hat{x}_a(t')-B_a(t,t'))+$$
$$-\frac{\Delta}{2}\sum_{it}\Sigma_i\hat{x}_i(t)\hat{x}_i(t)-\frac{\Delta}{2}\sum_{at}\Sigma_a\hat{x}_a(t)\hat{x}_a(t) \tag{3.2.12}$$

The first two lines and the last give $\mathcal{H}_\alpha$, the Hamiltonian for the interacting dynamics (3.2.1) but with the interaction terms $\phi_i(\boldsymbol{x}^{\mathrm{b}}(t),\boldsymbol{x}^{\mathrm{s}}(t))$ and $\phi_a(\boldsymbol{x}^{\mathrm{b}}(t),\boldsymbol{x}^{\mathrm{s}}(t))$ scaled by α. By definition of the Legendre transform, one can derive the conjugate fields as

$$\boldsymbol{h}^{\cdot}_{\alpha}=-\frac{1}{\Delta^n}\frac{\partial G_\alpha}{\partial \boldsymbol{m}^{\cdot}} \qquad \cdot=\mathrm{s},\mathrm{b} \tag{3.2.13}$$

where as in (3.2.12) we choose $n=1$ for linear fields and $n=2$ for quadratic ones to obtain well-defined values in the continuous time limit $\Delta\to 0$. The fields have components

$$\boldsymbol{h}^{\mathrm{b}}_{\alpha}=\{\boldsymbol{\Psi}^{\mathrm{b}}_{\alpha},\boldsymbol{l}^{\mathrm{b}}_{\alpha},\hat{\boldsymbol{C}}^{\mathrm{bb}}_{\alpha},\hat{\boldsymbol{R}}^{\mathrm{bb}}_{\alpha},\hat{\boldsymbol{B}}^{\mathrm{bb}}_{\alpha}\} \tag{3.2.14a}$$
$$\boldsymbol{h}^{\mathrm{s}}_{\alpha}=\{\boldsymbol{l}^{\mathrm{s}}_{\alpha},\hat{\boldsymbol{B}}^{\mathrm{ss}}_{\alpha}\} \tag{3.2.14b}$$

that are given explicitly by

$$\psi_{i\alpha}(t)=-\frac{1}{\Delta}\frac{\partial G_\alpha}{\partial \mu_i(t)} \tag{3.2.15a}$$
$$-\mathrm{i}l_{i\alpha}(t)=-\frac{1}{\Delta}\frac{\partial G_\alpha}{\partial(\hat{\mu}_i(t))} \tag{3.2.15b}$$
$$-\mathrm{i}l_{a\alpha}(t)=-\frac{1}{\Delta}\frac{\partial G_\alpha}{\partial(\hat{\mu}_a(t))} \tag{3.2.15c}$$
$$\hat{R}_{i\alpha}(t,t')=-\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial R_i(t',t)} \tag{3.2.15d}$$
$$\hat{C}_{i\alpha}(t,t')=-\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial C_i(t,t')} \tag{3.2.15e}$$
$$\hat{B}_{i\alpha}(t,t')=-\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial B_i(t,t')} \tag{3.2.15f}$$
$$\hat{B}_{a\alpha}(t,t')=-\frac{1}{\Delta^2}\frac{\partial G_\alpha}{\partial B_a(t,t')} \tag{3.2.15g}$$

The original dynamics has no biasing fields so the condition that defines the physical values of the order parameters is simply

$$\boldsymbol{h}^{\cdot}_{\alpha}=0 \qquad \cdot=\mathrm{s},\mathrm{b} \tag{3.2.16}$$

If one now Taylor expands to second order in α as $G_\alpha=G^0+\alpha G^1+(\alpha^2/2)G^2$, and similarly for $\boldsymbol{h}_\alpha$, then one sees that the physical order parameter values are obtained in the *non-interacting* ($\alpha=0$) theory by applying effective fields given by

$$\boldsymbol{h}^{\mathrm{eff}\,\cdot}=-\alpha\boldsymbol{h}^{1\,\cdot}-\frac{\alpha^2}{2}\boldsymbol{h}^{2\,\cdot} \qquad \cdot=\mathrm{s},\mathrm{b} \tag{3.2.17}$$

One has from (3.2.11)

$$\frac{d\Xi_\alpha}{d\alpha}=\frac{\partial\mathcal{H}_\alpha}{\partial\alpha}+\frac{\partial\boldsymbol{h}^{\mathrm{b}}_{\alpha}}{\partial\alpha}(\hat{\boldsymbol{m}}^{\mathrm{b}}-\boldsymbol{m}^{\mathrm{b}})+\frac{\partial\boldsymbol{h}^{\mathrm{s}}_{\alpha}}{\partial\alpha}(\hat{\boldsymbol{m}}^{\mathrm{s}}-\boldsymbol{m}^{\mathrm{s}}) \tag{3.2.18}$$

where one defines $\partial_\alpha \boldsymbol{h}^{\cdot}_{\alpha} = \boldsymbol{h}^{1\cdot}$, with $\cdot = \mathrm{s}, \mathrm{b}$ as before. Inserting this into the general expression for G^1 (see chapter 2) gives

$$\begin{aligned} G^1 &= \left\langle \frac{d\Xi_\alpha}{d\alpha} \right\rangle_0 = \left\langle -\Delta\Big[\sum_{at} \mathrm{i}\hat{x}_a(t)\phi_a(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) + \sum_{it} \mathrm{i}\hat{x}_i(t)\phi_i(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) \Big] \right\rangle_0 \\ &= -\Delta\Big[\sum_{at} \mathrm{i}\hat{\mu}_a(t)\Big(\sum_b J_{ab} x_b(t) + \sum_j K_{aj}\mu_j(t) \Big) + \sum_{it} \mathrm{i}\hat{\mu}_i(t)\Big(\sum_j J_{ij}\mu_j(t) + \sum_a K_{ia} x_a(t) \Big) \Big] \end{aligned} \tag{3.2.19}$$

where the average of the factors multiplying $\boldsymbol{h}^{1\cdot}$ with $\cdot = \mathrm{s}, \mathrm{b}$ vanishes by definition (3.2.9). For G^2 we need

$$\delta\Big(\frac{d\Xi_\alpha}{d\alpha}\Big) = \frac{d\Xi_\alpha}{d\alpha} - \left\langle \frac{d\Xi_\alpha}{d\alpha} \right\rangle_0 = -\Delta\Big[\sum_{ajt} \mathrm{i}\delta\hat{x}_a(t) K_{aj} \delta x_j(t) + \sum_{ijt} \mathrm{i}\delta\hat{x}_i(t) J_{ij} \delta x_j(t) \Big] \tag{3.2.20}$$

Here we have used the expression for the fields $\boldsymbol{h}^{1\cdot}$ for this linear case where we have assumed that there are no self-interactions ($J_{aa} = J_{ii} = 0$), namely

$$\psi^1_i(t) = \sum_a \mathrm{i}\hat{\mu}_a(t) K_{ai} + \sum_j \mathrm{i}\hat{\mu}_j(t) J_{ji} \tag{3.2.21a}$$

$$l^1_i(t) = -\sum_j J_{ij}\mu_j(t) - \sum_a K_{ia} x_a(t) \tag{3.2.21b}$$

$$l^1_a(t) = -\sum_j K_{aj}\mu_j(t) - \sum_b K_{ab} x_b(t) \tag{3.2.21c}$$

$$\hat{R}_i(t,t') = \hat{C}_i(t,t') = \hat{B}_i(t,t') = \hat{B}_a(t,t') = 0 \tag{3.2.21d}$$

(These can be derived by applying (3.2.13) to G^1 (3.2.19).) Inserting these into (3.2.18) and simplifying, one finds (3.2.20), where all terms containing non-fluctuating quantities (i.e. observations) have vanished. G^2 can now be calculated (see chapter 2) as

$$\begin{aligned} G^2 &= \left\langle \Big(\delta\Big(\frac{d\Xi_\alpha}{d\alpha} \Big) \Big)^2 \right\rangle_0 = \\ &= \Delta^2\Big[\sum_{ajtt'} K^2_{aj} \delta B_a(t,t') \delta C_j(t,t') + \sum_{ijtt'} J^2_{ij} \delta B_i(t,t') \delta C_j(t,t') + \sum_{ijtt'} J_{ij} J_{ji} \delta R_i(t,t') \delta R_j(t',t) \Big] \end{aligned} \tag{3.2.22}$$

where $\delta C_i(t,t')$, $\delta B_{a/i}(t,t')$ and $\delta R_i(t,t')$ denote the connected correlators and responses, e.g. $\delta C_i(t,t') = C_i(t,t') - \mu_i(t)\mu_i(t')$. All averages of products decouple into averages of variables at different nodes as $J_{aa} = J_{ii} = 0$. In addition, when the square is calculated, some of the contractions among terms are zero because of the factorization among different nodes in the effective dynamics at $\alpha = 0$. The last term, which contains anti-causal responses, must be retained at this level as it contributes to the derivatives (3.2.15) defining the fields. By taking the derivatives of G^2 we can find the second order contribution to the effective fields (3.2.17); combining with the first order terms above gives

(at $\alpha = 1$)

$$l_i^{\text{eff}}(t) = \sum_j J_{ij}\mu_j(t) + \sum_a K_{ia}x_a(t) + \int_0^\infty dt' \sum_j \mathrm{i}\hat{\mu}_i(t')J_{ij}^2\delta C_j(t,t') - \int_0^t dt' \sum_j J_{ij}J_{ji}\mu_i(t')\delta R_j(t',t) \tag{3.2.23a}$$

$$l_a^{\text{eff}}(t) = \sum_j K_{aj}\mu_j(t) + \sum_b K_{ab}x_b(t) + \int_0^\infty dt' \sum_a \mathrm{i}\hat{\mu}_a(t')K_{aj}^2\delta C_j(t,t') \tag{3.2.23b}$$

$$\psi_i^{\text{eff}}(t) = -\sum_a \mathrm{i}\hat{\mu}_a(t)K_{ai} - \sum_j \mathrm{i}\hat{\mu}_j(t)J_{ji} + \int_0^t dt' \sum_j J_{ij}J_{ji}\delta R_j(t',t)\mathrm{i}\hat{\mu}_i(t') - \int_0^\infty dt'\Big(\sum_a K_{ai}^2\delta B_a(t,t')\mu_i(t') + \sum_j J_{ij}^2\delta B_j(t,t')\mu_i(t')\Big) \tag{3.2.23c}$$

$$\hat{R}_i^{\text{eff}}(t',t) = \sum_j J_{ij}J_{ji}\delta R_j(t',t) \tag{3.2.23d}$$

$$\hat{C}_i^{\text{eff}}(t,t') = \sum_a K_{ai}^2\delta B_a(t,t') + \sum_j J_{ji}^2\delta B_j(t,t') \tag{3.2.23e}$$

$$\hat{B}_i^{\text{eff}}(t,t') = \sum_j J_{ij}^2\delta C_j(t,t') \tag{3.2.23f}$$

$$\hat{B}_a^{\text{eff}}(t,t') = \sum_j K_{aj}^2\delta C_j(t,t') \tag{3.2.23g}$$

These can be substituted into Ξ_0 to give the Plefka approximation for our partition function

$$P^{\text{eff}}(\boldsymbol{x}^{\text{s}}) = \int D\boldsymbol{x}^{\text{b}}D\hat{\boldsymbol{x}}^{\text{b}}D\hat{\boldsymbol{x}}^{\text{s}}\, \mathrm{e}^{\mathcal{H}^{\text{eff}}} \tag{3.2.24}$$

with

$$\begin{aligned}\mathcal{H}^{\text{eff}} \quad &= \sum_{at} \mathrm{i}\hat{x}_a(t)(x_a(t+\Delta) - x_a(t) + \Delta\lambda x_a(t) - \Delta l_a^{\text{eff}}(t)) - \frac{\Delta^2}{2}\sum_{att'}\Big(\hat{B}_a^{\text{eff}}(t,t') + \frac{\Sigma_a}{\Delta}\delta_{tt'}\Big)\hat{x}_a(t)\hat{x}_a(t') \\ &+ \sum_{it} \mathrm{i}\hat{x}_i(t)(x_i(t+\Delta) - x_i(t) + \Delta\lambda x_i(t) - \Delta l_i^{\text{eff}}(t) - \Delta^2\sum_{t'}\hat{R}_i^{\text{eff}}(t,t')x_i(t')) \\ &- \frac{\Delta^2}{2}\sum_{itt'}\Big(\hat{B}_i^{\text{eff}}(t,t') + \frac{\Sigma_i}{\Delta}\delta_{tt'}\Big)\hat{x}_i(t)\hat{x}_i(t') + \Delta\sum_{it}\psi_i^{\text{eff}}(t)x_i(t) + \frac{\Delta^2}{2}\sum_{itt'}\hat{C}_i^{\text{eff}}(t,t')x_i(t)x_i(t')\end{aligned} \tag{3.2.25}$$

The approximating action $\mathcal{H}^{\text{eff}}$ is factorized over sites, as anticipated above, thus the dynamics resulting from the action of the effective fields is non-interacting. As (3.2.25) is quadratic, it gives a Gaussian weight for the trajectories at each site.

Expression (3.2.24) then shows how the effective fields feature in the approximate posterior (conditional on the observations) dynamics. For the hidden nodes $x_i(t)$, similarly to chapter 2, one sees that $l_i^{\text{eff}}(t)$ is an effective drift; an additional coloured Gaussian noise also acts on $x_i(t)$, with covariance $\hat{B}_i^{\text{eff}}(t,t')$. Finally $\hat{R}_i^{\text{eff}}(t,t')$, as given by (3.2.23d), is a memory kernel with a simple intuitive interpretation: a fluctuation δx_i at time t' acts via J_{ji} as an effective field on x_j; at time

t this produces a response in x_j modulated by $R_j(t,t')$, which then acts back on x_i via J_{ij}. The presence of this term might seem to disagree with the dynamical mean field of asymmetric spin networks obtained by Kappen and Spanjers [72] and Mezard and Sakellariou [73]. As shown by Romano et al. [37], this apparent contradiction is solved, thus the Extended Plefka Expansion gives back those known results, once one considers the limit of large N^{b}, in which the memory with coefficient $J_{ij}J_{ji}$ can be neglected for asymmetric couplings and a J_{ij}^2 term is left from the coloured noise (3.2.23f). In addition, we note that also the observed dynamics acquires effective fields, with $l_a^{\mathrm{eff}}(t)$ a linear drift and $\hat{B}_a^{\mathrm{eff}}(t,t')$ the correlator of a Gaussian coloured noise on $x_a(t)$. Finally, the fields $\psi_i^{\mathrm{eff}}(t)$ and $\hat{C}_i^{\mathrm{eff}}(t,t')$, which would be zero without observations (see chapter 2) effectively constrain the hidden dynamics to be consistent with the observed trajectories, as we will explicitly show in the next section.

As result of including first and second moments (i.e. means, responses and correlations), the statistics of hidden (physical and auxiliary) and observed (auxiliary) trajectories is Gaussian within our approximation: this is known to be exact for a linear dynamics, except for the site-diagonal structure of second moments. The posterior means of the hidden variables can be regarded as the best estimate of the hidden dynamics, while the equal time posterior variance quantifies the degree of uncertainty for those inferred values and quantifies the inference error. Both the means and variances can be read off from the integrand in (3.2.24), which is proportional to the approximate Gaussian posterior over hidden trajectories conditioned on observations. The effective fields that couple to linear observables $(\psi_i^{\mathrm{eff}}(t), l_a^{\mathrm{eff}}(t), l_i^{\mathrm{eff}}(t))$ determine the posterior means; the ones that correspond to quadratic observables $(\hat{C}_i^{\mathrm{eff}}(t,t'), \hat{R}_i^{\mathrm{eff}}(t,t'), \hat{B}_i^{\mathrm{eff}}(t,t'), \hat{B}_a^{\mathrm{eff}}(t,t'))$ characterize the posterior second moments, i.e. posterior responses and correlations.

3.2.1 Posterior Means

For a Gaussian distribution in general the mean is the point where the probability density is stationary. To find the posterior mean we therefore just have to set the derivatives of the effective action to zero

$$\frac{\partial \mathcal{H}^{\mathrm{eff}}}{\partial \hat{x}_i(t)} = 0, \qquad \frac{\partial \mathcal{H}^{\mathrm{eff}}}{\partial x_i(t)} = 0 \tag{3.2.26}$$

One finds from these conditions (already in the continuous limit $\Delta \to 0$)

$$\begin{aligned} \partial_t \mu_i(t) &= -\lambda \mu_i(t) + l_i^{\mathrm{eff}}(t) + \int_0^t dt' \hat{R}_i^{\mathrm{eff}}(t,t')\mu_i(t') - \int_0^T dt' \left(\hat{B}_i^{\mathrm{eff}}(t,t') + \Sigma_i \delta(t-t') \right) \mathrm{i}\hat{\mu}_i(t') \\ &= -\lambda \mu_i(t) + \sum_j J_{ij}\mu_j(t) + \sum_a K_{ia} x_a(t) - \Sigma_i \mathrm{i}\hat{\mu}_i(t) \end{aligned} \tag{3.2.27}$$

which is the exact dynamics for the bulk plus an additional term with $\hat{\mu}_i(t)$ from conditioning on observations, which can also be viewed as a nonzero mean of the noise caused by conditioning. Another interpretation of $\hat{\mu}_i(t)$ is as a back-propagating error, as can be seen from its dynamics. The latter follows from the second part of (3.2.26) as

$$\begin{aligned}\partial_t(\mathrm{i}\hat{\mu}_i(t)) &= \lambda(\mathrm{i}\hat{\mu}_i(t)) + \psi_i^{\mathrm{eff}}(t) - \int_t^T dt' \hat{R}_i^{\mathrm{eff}}(t',t)(\mathrm{i}\hat{\mu}_i(t')) + \int_0^T dt' \hat{C}_i^{\mathrm{eff}}(t,t')\mu_i(t') = \\ &= \lambda(\mathrm{i}\hat{\mu}_i(t)) - \sum_a \mathrm{i}\hat{\mu}_a(t)K_{ai} - \sum_j \mathrm{i}\hat{\mu}_j(t)J_{ji} \end{aligned} \tag{3.2.28}$$

We can now more clearly understand the roles played by the effective fields in the auxiliary dynamics, with $\psi_i^{\mathrm{eff}}(t)$ being an effective drift and $\hat{C}_i^{\mathrm{eff}}(t,t')$ the covariance of an additional coloured noise. The integration over $T > t' > t$ indicates that $\hat{R}_i^{\mathrm{eff}}(t,t')$ can also be interpreted as a memory kernel in this context, but it is a memory "from the future" in the sense that it provides the weight with which the future values of $\hat{\mu}_i(t)$ affect the present ones. The dynamics it-self of $\hat{\mu}_i(t)$ propagates its values backwards from the final time T. This dynamical evolution depends on the $\hat{\mu}_a(t)$, which in turn *does depend* on the dynamics of observations: in this way, observations make auxiliary variables non-zero. In fact one has from the analogue for observations of conditions (3.2.26), i.e. $\partial\mathcal{H}^{\mathrm{eff}}/\partial\hat{x}_a(t) = 0$,

$$\begin{aligned}\partial_t x_a(t) &= -\lambda x_a(t) + I_a^{\mathrm{eff}}(t) - \int_0^T dt' \Big(\hat{B}_a^{\mathrm{eff}}(t,t') + \Sigma_a\delta(t-t')\Big)\mathrm{i}\hat{\mu}_a(t') = \\ &= -\lambda x_a(t) + \sum_j K_{aj}\mu_j(t) + \sum_b K_{ab}x_b(t) - \Sigma_a\mathrm{i}\hat{\mu}_a(t) \end{aligned} \tag{3.2.29}$$

To better grasp how the conditioning on observations enters the dynamics by a backward propagation, we shall briefly go back to the discrete time version of equations (3.2.27), (3.2.28), (3.2.29)

$$\frac{\mu_i(t+\Delta) - \mu_i(t)}{\Delta} = -\lambda\mu_i(t) + \sum_j J_{ij}\mu_j(t) + \sum_a K_{ia}x_a(t) - \Sigma_i\mathrm{i}\hat{\mu}_i(t) \tag{3.2.30}$$

$$\frac{\mathrm{i}\hat{\mu}_i(t) - \mathrm{i}\hat{\mu}_i(t-\Delta)}{\Delta} = \lambda(\mathrm{i}\hat{\mu}_i(t)) - \sum_a \mathrm{i}\hat{\mu}_a(t)K_{ai} - \sum_j \mathrm{i}\hat{\mu}_j(t)J_{ji} \tag{3.2.31}$$

$$\frac{x_a(t+\Delta) - x_a(t)}{\Delta} = -\lambda x_a(t) + \sum_j K_{aj}\mu_j(t) + \sum_b K_{ab}x_b(t) - \Sigma_a\mathrm{i}\hat{\mu}_a(t) \tag{3.2.32}$$

Here Δ is the unit time step and the time index $t = \Delta, ..., T$ for $\{\mu_i(t)\}$ while $t = 0, ..., T-\Delta$ for $\{\hat{\mu}_i(t)\}$ and $\{\hat{\mu}_a(t)\}$: therefore, *by construction* of the path integral representation of the dynamics, the stationarity conditions give $\hat{\mu}_i(T) = \hat{\mu}_a(T) = 0$ (see also for more details a general review on path integral methods [23]). The "initial" condition $\hat{\mu}_a(T-\Delta)$ is determined by the observed values at the end of the trajectory $x_a(T)$, $x_a(T-\Delta)$, as is clear from (3.2.32). Equation (3.2.31) calculated at $t = T$ and $t = T-\Delta$ shows that $\hat{\mu}_i(T-\Delta)$ vanishes but $\hat{\mu}_i(T-2\Delta)$ assumes a value different from zero because of $\hat{\mu}_a(T-\Delta)$; in particular $\hat{\mu}_i(T-2\Delta) = \Delta\sum_a \hat{\mu}_a(T-\Delta)K_{ai}$. This

closer inspection time step by time step immediately reveals that the fixed values of observations introduce a non-zero correction term $\hat{\mu}_i(t)$ in the average bulk dynamics (3.2.30) which evolves backward in time (i.e. the "initial" values, which are zero, are at $t = T - \Delta$).

Together, the above equations for the posterior means (3.2.27), (3.2.28), (3.2.29) constitute a forward-backward propagation, with the $\hat{\mu}_i$ responsible for the backward portion, i.e. for the flow of information from the future. The forward-backward structure is consistent with the general theory of conditional stochastic processes, for which the estimation of posterior distributions requires information to propagate both in the forward and backward direction [74].

To summarize, the averages of auxiliary variables do not vanish identically as in the Extended Plefka Expansion without observations (see chapter 2) where the mean (i.e. stationary) path of a linear dynamics would be given by a noiseless relaxation. The observed trajectories effectively "bias" the path ensemble in such a way that the mean hidden conditional dynamics follows what would be called an *activation* path (see e.g. [75]), i.e. for which auxiliary variables are different from zero: these can be therefore seen as implementing the constraints given by the observations. This is compatible with the fact that the path integral (3.2.4) is not expected to be equal to one since it is interpreted as the data likelihood rather than the normalization of a probability distribution.

In addition, one can show that (3.2.27), (3.2.28), (3.2.29) map exactly onto the time evolution for means and fields obtained in [59] by saddle point approximation, provided that one considers small couplings so that the hyperbolic tangents in the mean spin dynamics of [59] can be linearized. In fact, in this linear dynamics we study, (3.2.27), (3.2.28), (3.2.29) can be obtained by saddle point method starting from the exact action (3.2.5) rather than the approximated Plefka one (3.2.25) (whose application is of course more general than the linear case).

Finally we stress that the combination of (3.2.27), (3.2.28), (3.2.29) gives, as prediction for the conditional dynamics, the mean of a Gaussian path integral, thus corresponds to the optimal solution as computable from the Kalman filter recursion [16, 63]. We summarize some results from the Kalman filter relevant to our analysis in appendix E and for a direct comparison with equations (3.2.27), (3.2.28), (3.2.29) see appendix I, section I.1.1.

3.2.2 Posterior Variance

The inverse covariance of the approximating Gaussian distribution can be read off from the path integral representation of the likelihood (3.2.24). In particular it consists of two distinct blocks, as the distribution is factorized w.r.t. the subnetwork s and bulk b indices. These blocks can be

written symbolically as

$$\boldsymbol{C}^{-1}_{i\,\mathrm{gen}}(t,t') = \begin{pmatrix} -\hat{C}^{\mathrm{eff}}_i(t,t') & (\partial_{t'}+\lambda)\delta(t-t') - \hat{R}^{\mathrm{eff}}_i(t',t) \\ (\partial_t+\lambda)\delta(t-t') - \hat{R}^{\mathrm{eff}}_i(t,t') & -\hat{B}^{\mathrm{eff}}_i(t,t') - \Sigma_i\delta(t-t') \end{pmatrix} \tag{3.2.33}$$

$$C^{-1}_{a\,\mathrm{gen}}(t,t') = -\hat{B}^{\mathrm{eff}}_a(t,t') - \Sigma_a\delta(t-t') \tag{3.2.34}$$

The "gen" subscript indicates a "generalized" inverse covariance including the auxiliary variables $\hat{x}_i(t)$, $\hat{x}_a(t)$, i.e.

$$\boldsymbol{C}_{i\,\mathrm{gen}}(t,t') = \left\langle \begin{pmatrix} \delta x_i(t) \\ -\mathrm{i}\delta\hat{x}_i(t) \end{pmatrix} \begin{pmatrix} \delta x_i(t') & -\mathrm{i}\delta\hat{x}_i(t') \end{pmatrix} \right\rangle$$

while $C_{a\,\mathrm{gen}}(t,t') = -\langle \delta\hat{x}_a(t)\delta\hat{x}_a(t')\rangle$ refers *only* to the auxiliary variables $\hat{x}_a(t)$ as the $x_a(t)$ are fixed by the observations (here $\delta x_i(t)$ indicates the deviation from the mean, i.e. $\delta x_i(t) = x_i(t) - \mu_i(t)$ and similarly for $\delta\hat{x}_i(t)$ and $\delta\hat{x}_a(t)$). In principle all of the second order functions are connected but in what follows we drop the δs for the sake of brevity. The covariance itself then has the same block structure

$$\boldsymbol{C}_{i\,\mathrm{gen}}(t,t') = \begin{pmatrix} C_i(t,t') & R_i(t,t') \\ R_i(t',t) & B_i(t,t') \end{pmatrix} \tag{3.2.35}$$

$$C_{a\,\mathrm{gen}}(t,t') = B_a(t,t') \tag{3.2.36}$$

Note that here and below we write directly the continuous time equations that are obtained from the discrete time formalism as $\Delta \to 0$ (see appendix D for more details). To find the equations for the (generalized) covariances, one just takes the identity $\int d\tau\, \mathbf{C}_{i\,\mathrm{gen}}(t,\tau)\mathbf{C}^{-1}_{i\,\mathrm{gen}}(\tau,t') = \mathbb{1}\delta(t-t')$ block by block to get

$$\int_{-\infty}^{+\infty} d\tau \Big[-C_i(t,\tau)\hat{C}^{\mathrm{eff}}_i(\tau,t') + R_i(t,\tau)\Big((\partial_\tau+\lambda)\delta(\tau-t') - \hat{R}^{\mathrm{eff}}_i(\tau,t')\Big)\Big] = \delta(t-t') \tag{3.2.37a}$$

$$\int_{-\infty}^{+\infty} d\tau \Big[C_i(t,\tau)\Big((\partial_{t'}+\lambda)\delta(t'-\tau) - \hat{R}^{\mathrm{eff}}_i(t',\tau)\Big) - R_i(t,\tau)\Big(\hat{B}^{\mathrm{eff}}_i(\tau,t') + \Sigma_i\delta(\tau-t')\Big)\Big] = 0 \tag{3.2.37b}$$

$$\int_{-\infty}^{+\infty} d\tau \Big[R_i(\tau,t)\Big((\partial_{t'}+\lambda)\delta(t'-\tau) - \hat{R}^{\mathrm{eff}}_i(t',\tau)\Big) - B_i(t,\tau)\Big(\hat{B}^{\mathrm{eff}}_i(\tau,t') + \Sigma_i\delta(\tau-t')\Big)\Big] = \delta(t-t') \tag{3.2.37c}$$

and similarly for the auxiliary variables

$$\int_{-\infty}^{+\infty} d\tau\, C_{a\,\mathrm{gen}}(t,\tau)C^{-1}_{a\,\mathrm{gen}}(\tau,t') = -\int_{-\infty}^{+\infty} d\tau\, B_a(t,\tau)\Big[\hat{B}^{\mathrm{eff}}_a(\tau,t') + \Sigma_a\delta(\tau-t')\Big] = \delta(t-t') \tag{3.2.38}$$

If we now substitute the expressions for $\hat{R}^{\mathrm{eff}}_i(t',t)$, $\hat{C}^{\mathrm{eff}}_i(t,t')$, $\hat{B}^{\mathrm{eff}}_i(t,t')$ and $\hat{B}^{\mathrm{eff}}_a(t,t')$ from (3.2.23) we get a closed system of integral equations. To simplify these, we consider long times, where a stationary regime is reached so that all two-time functions become Time Translation Invariant (TTI). In an inference problem, stationarity implies also that the quality of the prediction is the same at all times. At stationarity, the integrals in the above equations (3.2.37) and (3.2.38) become

convolutions and hence, if we go to double-sided Laplace transform, simple products. After some simplification the Laplace-transformed equations yield a system of four coupled equations for $\tilde{C}_i(z)$, $\tilde{R}_i(z)$, $\tilde{B}_i(z)$, $\tilde{B}_a(z)$

$$-\tilde{C}_i(z)\Big(\sum_a K_{ai}^2\tilde{B}_a(z)+\sum_j J_{ji}^2\tilde{B}_j(z)\Big)+\tilde{R}_i(z)\Big(z+\lambda-\sum_j J_{ij}J_{ji}\tilde{R}_j(z)\Big)=1 \tag{3.2.39a}$$

$$\tilde{C}_i(z)\Big(-z+\lambda-\sum_j J_{ij}J_{ji}\tilde{R}_j(-z)\Big)-\tilde{R}_i(z)\Big(\sum_j J_{ij}^2\tilde{C}_j(z)+\Sigma_i\Big)=0 \tag{3.2.39b}$$

$$\begin{aligned}\tilde{B}_i(z)\Big[\Big(-z+\lambda-\sum_j J_{ij}J_{ji}\tilde{R}_j(-z)\Big)\Big(\sum_a K_{ai}^2\tilde{B}_a(z)+\sum_j J_{ji}^2\tilde{B}_j(z)\Big)^{-1}\\ \Big(z+\lambda-\sum_j J_{ij}J_{ji}\tilde{R}_j(z)\Big)-\Big(\sum_j J_{ij}^2\tilde{C}_j(z)+\Sigma_i\Big)\Big]=1\end{aligned} \tag{3.2.39c}$$

$$\tilde{B}_a(z)\Big[\Sigma_a+\sum_j K_{aj}^2\tilde{C}_j(z)\Big]=-1 \tag{3.2.39d}$$

In (3.2.39c) we have substituted the expression for $\tilde{R}_i(z)$ as given by (3.2.39b) and used the fact that $\tilde{C}_i(z)\Big(\sum_j J_{ij}^2\tilde{C}_j(z)+\Sigma_i\Big)^{-1}=\tilde{B}_i(z)\Big(\sum_a K_{ai}^2\tilde{B}_a(z)+\sum_j J_{ji}^2\tilde{B}_j(z)\Big)^{-1}$, as can be deduced by comparison of (3.2.37a) and (3.2.37c).

3.2.3 Thermodynamic Limit

We expect the extended Plefka approach to give exact values for posterior means and variances in the case of mean field type couplings (i.e. weak and long-ranged), in the thermodynamic limit of an infinitely large system. More precisely we define the thermodynamic limit as the one of an infinitely large bulk and subnetwork, $N^{\mathrm{b}}, N^{\mathrm{s}} \to \infty$ at constant ratio $\alpha = N^{\mathrm{s}}/N^{\mathrm{b}}$. Note that in the following discussion, α indicates this ratio and not the Plefka perturbative parameter, which in all these results is set to 1. For our mean field couplings we assume, as in chapter 2, that $\{J_{ij}\}$ is a real matrix belonging to the Girko ensemble [43], i.e. its elements are independently and randomly distributed Gaussian variables with zero mean and variance satisfying

$$\langle J_{ij}J_{ij}\rangle = \frac{j^2}{N^{\mathrm{b}}} \tag{3.2.40}$$

$$\langle J_{ji}J_{ij}\rangle = \frac{\eta\, j^2}{N^{\mathrm{b}}} \tag{3.2.41}$$

The parameter $\eta \in [-1,1]$ controls the degree to which the matrix $\{J_{ij}\}$ is symmetric. In particular, the dynamics is non-equilibrium – it does not satisfy detailed balance – whenever $\eta < 1$. Similarly, $\{K_{ai}\}$ is taken as a random matrix with uncorrelated zero mean Gaussian entries of variance

$$\langle K_{ai}\rangle = 0 \qquad \langle K_{ai}^2\rangle = \frac{k^2}{N^{\mathrm{b}}} \tag{3.2.42}$$

We have introduced amplitude parameters j for $\{J_{ij}\}$ (hidden-to-hidden) and k for $\{K_{ai}\}$ (hidden-to-observed) here. The scaling of both types of interaction parameters with $1/\sqrt{N^{\mathrm{b}}}$ ensures that, when the size of the hidden part of the system increases, the typical contribution it makes to the time evolution of each hidden and observed variable stays of the same order. The high connectivity, where all nodes interact with all other ones, implies that in the thermodynamic limit all nodes become equivalent. Local response and correlation functions therefore become identical to their averages over nodes, defined as

$$\tilde{R}^{\mathrm{bb|s}}(z) = \frac{1}{N^{\mathrm{b}}} \sum_j \tilde{R}_j(z) \tag{3.2.43}$$

$$\tilde{C}^{\mathrm{bb|s}}(z) = \frac{1}{N^{\mathrm{b}}} \sum_j \tilde{C}_j(z) \tag{3.2.44}$$

$$\tilde{B}^{\mathrm{bb|s}}(z) = \frac{1}{N^{\mathrm{b}}} \sum_j \tilde{B}_j(z) \tag{3.2.45}$$

As a consequence, all site indices can be dropped and the correlation and response functions can be replaced by their mean values, $\tilde{C}_i(z) \equiv \tilde{C}^{\mathrm{bb|s}}(z)$, $\tilde{R}_i(z) \equiv \tilde{R}^{\mathrm{bb|s}}(z)$, $\tilde{B}_i(z) \equiv \tilde{B}^{\mathrm{bb|s}}(z)$. The superscripts bb|s here emphasize that we are considering moments conditioned on subnetwork values though for brevity we drop them in the rest of the calculation. Similarly, for the correlations of auxiliary variables related to observations we can set $\tilde{B}_a(z) \equiv \tilde{B}^{\mathrm{ss}}(z)$. Of primary interest is then $\tilde{C}(z)$, the Laplace transformed posterior (co-)variance function of prediction errors, which as a site-average can be thought of as a macroscopic measure of prediction performance. This should become self-averaging in the thermodynamic limit. To see this, consider e.g. the sum $\sum_j J_{ij}J_{ji}\tilde{R}_j(z)$. Replacing $\tilde{R}_j$ by $\tilde{R}$, the prefactor is a sum of N^{b} terms so converges to its average $N^{\mathrm{b}}\eta j^2/N^{\mathrm{b}} = \eta j^2$ in the thermodynamic limit. So, as in chapter 2, we can replace

$$\sum_j J_{ij}J_{ji}\tilde{R}_j(z) \sim \eta\, j^2\tilde{R}(z)$$

Making this and similar substitutions in the system (3.2.39), and choosing scalar noise covariances $\Sigma_i = \sigma_{\mathrm{b}}^2$ and $\Sigma_a = \sigma_{\mathrm{s}}^2$ for simplicity, one finds

$$-\tilde{C}(z)\Big(-\frac{\alpha k^2}{\sigma_{\mathrm{s}}^2 + k^2\tilde{C}(z)} + j^2\tilde{B}(z)\Big) + \tilde{R}(z)\Big(z + \lambda - \eta j^2\tilde{R}(z)\Big) = 1 \tag{3.2.46a}$$

$$\tilde{C}(z)\Big(-z + \lambda - \eta j^2\tilde{R}(-z)\Big) - \tilde{R}(z)\Big(j^2\tilde{C}(z) + \sigma_{\mathrm{b}}^2\Big) = 0 \tag{3.2.46b}$$

$$\begin{aligned}\tilde{B}(z)\Big[\Big(-z + \lambda - \eta j^2\tilde{R}(-z)\Big)\Big(-\frac{\alpha k^2}{\sigma_{\mathrm{s}}^2 + k^2\tilde{C}(z)} + j^2\tilde{B}(z)\Big)^{-1}\Big(z + \lambda - \eta j^2\tilde{R}(z)\Big) \\ - \Big(j^2\tilde{C}(z) + \sigma_{\mathrm{b}}^2\Big)\Big] = 1\end{aligned} \tag{3.2.46c}$$

Here $\tilde{B}^{\mathrm{ss|s}}(z) = -1/(\sigma_{\mathrm{s}}^2 + k^2\tilde{C}(z))$ has already been substituted into equations (3.2.46a) and (3.2.46c).

We comment briefly on the relation of the above results to the Fluctuation Dissipation Theorem (FDT), which in terms of Laplace transforms reads

$$z\tilde{C}(z) = -\frac{\sigma_{\rm b}^2}{2}[\tilde{R}(z) - \tilde{R}(-z)] \tag{3.2.47}$$

This can be compared with the expression for $\tilde{R}(z)$ that follows from (3.2.46b) (taken for z and $-z$)

$$\tilde{R}(z) = \tilde{C}(z)\Big[\frac{\lambda}{\sigma_{\rm b}^2 + j^2(1+\eta)\tilde{C}(z)} - \frac{z}{j^2(1-\eta)\tilde{C}(z) + \sigma_{\rm b}^2}\Big] \tag{3.2.48}$$

The r.h.s. of the FDT is then

$$-\frac{\sigma_{\rm b}^2}{2}[\tilde{R}(z) - \tilde{R}(-z)] = z\tilde{C}(z)\frac{\sigma_{\rm b}^2}{j^2(1-\eta)\tilde{C}(z) + \sigma_{\rm b}^2} \tag{3.2.49}$$

Comparing with (3.2.47), the FDT is satisfied for symmetric couplings ($\eta = 1$) as expected, while there are progressively stronger deviations from FDT as η decreases towards -1.

3.2.4 Comparison with known results

As a consistency check and to support our claim of exactness in the thermodynamic limit, we briefly look at how $\tilde{C}(z)$ and $\tilde{R}(z)$ can be worked out by alternative means.

In general, from (3.2.46b) we can get an expression for $\tilde{R}(z)$ in terms of $\tilde{C}(z)$: this is (3.2.48). Substituting into (3.2.46a), $\tilde{B}(z)$ can also be worked out as a function of $\tilde{C}(z)$. Using these expressions for $\tilde{R}(z)$ and $\tilde{B}(z)$ in equation (3.2.46c), one gets a closed algebraic equation for the posterior variance $\tilde{C}(z)$

$$z^2 = \left[-\frac{\sigma_{\rm b}^2}{\tilde{C}} + \frac{\alpha\frac{k^2\sigma_{\rm b}^2}{\sigma_{\rm s}^2}}{1+\frac{k^2}{\sigma_{\rm s}^2}\tilde{C}} + \frac{j^2}{1+\frac{j^2}{\sigma_{\rm b}^2}\tilde{C}} + \frac{\lambda^2}{\left(1+(1+\eta)\frac{j^2}{\sigma_{\rm b}^2}\tilde{C}\right)^2}\right]\left(1+(1-\eta)\frac{j^2}{\sigma_{\rm b}^2}\tilde{C}\right)^2 \tag{3.2.50}$$

This is the same expression we will obtain in chapter 4 using an explicit average over the quenched disorder variables J_{ij} and K_{ia}. Particular cases are also further validated by random matrix theory, as follows.

Case $\alpha = 0$.

This case corresponds to the absence of observations. One has then $\tilde{B}_i(z)$, $\tilde{B}_a(z) \equiv 0$ as these quantities simply play the role of Lagrange multipliers enforcing the conditioning on observations. To see this formally from the $\alpha \to 0$ limit of (3.2.46) one sets $\tilde{B}_i(z) = \alpha\tilde{D}_i(z)$ and $\tilde{B}_a(z) = \alpha\tilde{D}_a(z)$ where the $\tilde{D}$s stay nonzero for $\alpha \to 0$. One verifies that under this assumption the system has as

solution the responses and correlations known from chapter 2

$$\tilde{R}(z) = \frac{1}{2\eta}(z+\lambda) - \frac{1}{2\eta}\sqrt{(z+\lambda)^2 - 4j^2\eta} \tag{3.2.51}$$

$$\tilde{C}(z) = \frac{4\,\sigma_{\mathrm{b}}^2}{[(\lambda+z) + \sqrt{(\lambda+z)^2 - 4j^2\eta}][(\lambda-z) + \sqrt{(\lambda-z)^2 - 4j^2\eta}] - 4j^2} \tag{3.2.52}$$

Case $j = 0$.

In this situation, hidden variables are uncoupled (i.e. coupled only via a self-interaction term λ). By solving (3.2.50), one obtains

$$\tilde{C}(z) = \frac{\sigma_{\mathrm{s}}^2}{2k^2}\frac{\sigma^2}{(\lambda^2 - z^2)}\left\{1 - \alpha - \left(\frac{\lambda^2 - z^2}{\sigma^2}\right) + \sqrt{\left[1 - \alpha - \left(\frac{\lambda^2 - z^2}{\sigma^2}\right)\right]^2 + 4\left(\frac{\lambda^2 - z^2}{\sigma^2}\right)}\right\} \tag{3.2.53}$$

where $\sigma = \sigma_{\mathrm{b}}k/\sigma_{\mathrm{s}}$. This coincides with the result we will derive in chapter 4 using methods from random matrix theory.

Case $\eta = 1$.

Here we have symmetric hidden-to-hidden couplings and (3.2.50) can be simplified to

$$z^2 = -\frac{\sigma_{\mathrm{b}}^2}{\tilde{C}(z)} + \frac{\alpha\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}}{1 + \frac{k^2}{\sigma_{\mathrm{s}}^2}\tilde{C}(z)} + \frac{j^2}{1 + \frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}(z)} + \frac{\lambda^2}{\left[1 + 2\frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}(z)\right]^2} \tag{3.2.54}$$

This fifth order equation for the single site posterior covariance predicted by the extended Plefka expansion with hidden nodes is again confirmed by random matrix theory results (see chapter 4).

3.3 Power spectrum and critical regions

In this section we study the predictions of the extended Plefka approach for our conditional dynamical system, focussing on the power spectrum of the posterior covariance. This is given by $\tilde{C}(\mathrm{i}\omega)$, the Laplace transform $\tilde{C}(z)$ evaluated at $z = \mathrm{i}\omega$.

3.3.1 Dimensionless system

We would like to understand how $\tilde{C}(\omega)$ depends on the parameters λ, j, k, σ_{s}, σ_{b}, η and α. The last two of these are already dimensionless. By extracting the appropriate dimensional scales, we can reduce the other five parameters to only two dimensionless combinations.

From (3.2.1) one sees that j, k, λ have dimensions t^{-1}, while the dimension of $\sigma^2_{\mathrm{s/b}}$ is x^2t^{-1}. We can build from these the dimensionless parameters $\gamma = j/\lambda$ and $p = \lambda/\sigma$. Here $\sigma = \sigma_{\mathrm{b}}k/\sigma_{\mathrm{s}}$, which has dimension t^{-1} and contains the observation "intensity" k as well as the ratio between the dynamical noises $\sigma_{\mathrm{b}}/\sigma_{\mathrm{s}}$. The latter is a third dimensionless parameter but as it only enters one prefactor we will not need to keep it separately.

Extracting appropriate dimensional amplitudes for all four two-point functions, we write them as

$$\tilde{C}(\mathrm{i}\omega) = \frac{\sigma_{\mathrm{s}}^2}{k^2} C_{\alpha,\gamma,\eta,p}(\Omega) \tag{3.3.1a}$$

$$\tilde{R}(\mathrm{i}\omega) = \frac{1}{j} \mathcal{R}_{\alpha,\gamma,\eta,p}(\Omega) \tag{3.3.1b}$$

$$\tilde{B}(\mathrm{i}\omega) = \frac{1}{\sigma_{\mathrm{b}}^2} \mathcal{B}_{\alpha,\gamma,\eta,p}(\Omega) \tag{3.3.1c}$$

$$\tilde{B}^{\mathrm{ss}}(\mathrm{i}\omega) = \frac{1}{\sigma_{\mathrm{s}}^2} \mathcal{B}^{\mathrm{ss}}_{\alpha,\gamma,\eta,p}(\Omega) \tag{3.3.1d}$$

Here $\Omega = \omega/\sigma$ is a dimensionless frequency; similarly $C_{\alpha,\gamma,\eta,p}(\Omega)$, $\mathcal{R}_{\alpha,\gamma,\eta,p}(\Omega)$, $\mathcal{B}_{\alpha,\gamma,\eta,p}(\Omega)$ and $\mathcal{B}^{\mathrm{ss}}_{\alpha,\gamma,\eta,p}(\Omega)$ are dimensionless and depend on the dimensionless parameters α, γ, η and p: for the sake of brevity, we do not write the subscripts indicating this dependence in the following.

Let us briefly comment on (3.3.1a). One sees that $\tilde{C}(z)$ is directly proportional to σ_{s}^2 and inversely proportional to k^2: the weaker the hidden-to-observed coupling and the stronger the dynamical noise acting on the observed variables, the less information one can extract from the subnetwork trajectories and the more uncertain the predictions for the behaviour of the bulk.

To summarize, we switch from eight original parameters $\{\alpha, \eta, \lambda, j, k, \sigma_{\mathrm{s}}, \sigma_{\mathrm{b}}, \omega\}$ to a set of five dimensionless parameters $\{\alpha, \eta, \gamma, p, \Omega\}$. For the dimensionless second moments (3.3.1), the system (3.2.46) becomes

$$-C\left(-\frac{\alpha}{1+C} + (\gamma p)^2\mathcal{B}\right) + (\gamma p)^{-1}\mathcal{R}\left(\mathrm{i}\Omega + p - \eta\gamma p\mathcal{R}\right) = 1 \tag{3.3.2a}$$

$$(\gamma p)\,C\left(-\mathrm{i}\Omega + p - \eta\gamma p\mathcal{R}_-\right) - \mathcal{R}\left((\gamma p)^2 C + 1\right) = 0 \tag{3.3.2b}$$

$$\mathcal{B}\left[\left(-\mathrm{i}\Omega + p - \eta\gamma p\mathcal{R}_-\right)\left(-\frac{\alpha}{1+C} + (\gamma p)^2\mathcal{B}\right)^{-1}\left(\mathrm{i}\Omega + p - \eta\gamma p\mathcal{R}\right) - (\gamma p)^2 C - 1\right] = 1 \tag{3.3.2c}$$

where we have dropped the frequency argument and introduced the shorthand $\mathcal{R}_- = \mathcal{R}(-\Omega)$.

3.3.2 Critical scaling

Scaling concepts are fundamental in Quantum Field Theory (applied to particle physics) and equilibrium analysis of criticality, as they lie at the interface between theory and experiment (see [76] for an exhaustive presentation). The main idea, borrowed from the Renormalization Group [76],

is to focus on some observables at particular length or time scales: postulating scaling relations between them allows one to remove singularities in their definition and to recover a regular behaviour. When only one independent variable can be identified, these relations assume the shape of power laws. Close to the critical point, where divergencies arise, we look at the system at a different scale and this change of reference frame is achieved by rescaling amplitude, frequency and other parameters ("fields" in the language of phase transitions) by some power of the distance from the critical point (which indeed goes to zero approaching it). The footing of this procedure consists first of working out the power laws that describe the dependence on such a quantity: the renormalization w.r.t. it produces expressions for the correlations that are invariant under that particular scale transformation. As we will explain in more detail, the arbitrariness in setting the scaling parameter can be used to reduce the number of independent variables by one.

We next analyze in the parameter space α, γ, p (at fixed η) the singularities of $C(0)$, the (dimensionless) zero frequency posterior covariance. The behaviour of the inference error itself, the equal time posterior variance $C(t-t) = C(0)$, will be addressed in section 3.4.3.

Independently of η, we find two critical regions that are shown graphically in figure 3.1:

1. $\forall p$, $\alpha = 0$ and $\gamma > \gamma_c$
2. $\forall \gamma$, $p = 0$ and $0 < \alpha < 1$

The first case gives back the dynamics without observations ($\alpha = 0$), for which $\gamma < \gamma_c = 1/(1+\eta)$ is the condition of stability beyond which trajectories typically diverge in time. Interestingly, as soon as $\alpha > 0$, the constraints from observations make the solution stable irrespective of whether γ is smaller or bigger than the critical value. For $\gamma > \gamma_c$ the observed trajectories would then be divergent, and so would the predicted hidden trajectories, while the error (posterior variance) of the predictions would remain bounded. It is difficult to conceive of situations where divergent mean trajectories would make sense, however, so we only consider the range $\gamma \leq \gamma_c$ in our analysis.

The second limit, $p \to 0$, corresponds, for fixed ratio between noises σ_s/σ_b, to $k \gg \lambda$: we call this scenario an "underconstrained" hidden system. In general for large k the posterior variance decreases as $1/k^2$, as used in the scaling (3.3.1a). But for $\alpha < 1$ there are directions in the space of hidden trajectories that are not constrained at all by subnetwork observations, and their variance will scale as $1/\lambda^2$ instead. These directions give a large contributions to the dimensionless $C(\Omega)$ that diverges in the limit $k \gg \lambda$. In general one has a similar effect when $k/\sigma_s \gg \lambda/\sigma_b$, where the noise in the dynamics acts to effectively reduce the relevant interaction or decay constant. This behaviour is broadly analogous to what happens in learning of linear functions from static data

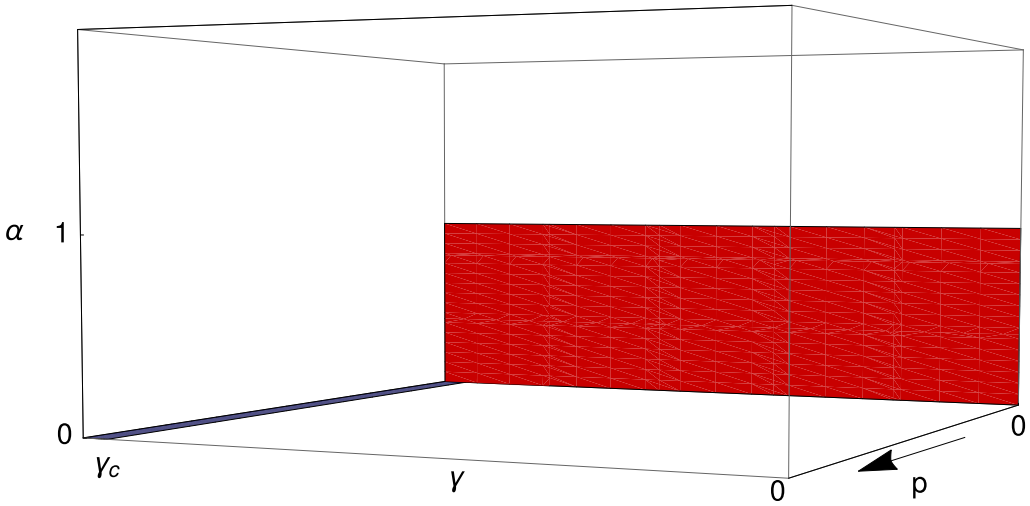

Figure 3.1: Parameter space spanned by α, γ, p. The blue stripe and the red area mark the values for which the posterior covariance $C(0)$ becomes singular, i.e. respectively $\alpha = 0$, $\gamma > \gamma_c$ $(\forall p)$ and $p = 0$, $0 < \alpha < 1$ $(\forall \gamma)$.

[65, 66]: there the prediction error will also diverge when α, which is then defined as the ratio of number of examples to number of spatial dimensions, is less than unity and no regularization is applied.

An important point to stress is that these critical regions do not arise as features of the Plefka expansion to any extent. It is well known that the Plefka expansion has a finite radius of convergence [33] in terms of couplings strength, thus in principle it should be checked whether our parameters lie in the region of validity of the expansion or not. The agreement of our results (3.3.2) with other methods which do not rely on perturbative expansions (random matrix theory and functional average over the disorder, see chapter 4) clearly suggests that there is no problem of convergence where the system is well defined.

Close to the two critical regions in the space of α, γ and p, $C(0)$ is expected to exhibit a power-law dependence on the parameters specifying the distance away from the critical point. We can then perform a systematic rescaling analysis, as is standard in the study of phase transitions [76]. The aim of this analysis, developed in the next two sections, is to find the master curves and associated exponents that describe the approach to the critical point(s).

3.3.3 Master curves for $\gamma \to \gamma_c$ and $\alpha \to 0$

We begin with the behaviour of the posterior covariance power spectrum for $\gamma \to \gamma_c$ and $\alpha \to 0$, i.e. in the vicinity of the "no observations" critical point.

We already know the limit curve for the power spectrum: it is given by the spectrum *at* the critical point. At this point we have an $\alpha = 0$ system, i.e. without observations (see chapter 2 and the summary in section 3.2.4). As there the interactions and noise level relating to observed nodes are then irrelevant, we can set $k = j$, $\sigma_b = \sigma_s$ (i.e $p = 1/\gamma$, $\sigma = j$ and $\Omega = \omega/j$). From (3.2.52) the dimensionless correlation is then

$$C_{\gamma,\eta}(\Omega) = \frac{4}{\left[\frac{1}{\gamma} + i\Omega + \sqrt{(\frac{1}{\gamma} + i\Omega)^2 - 4\eta}\right]\left[\frac{1}{\gamma} - i\Omega + \sqrt{(\frac{1}{\gamma} - i\Omega)^2 - 4\eta}\right] - \frac{4}{\sigma^2}} \tag{3.3.3}$$

$C(0)$ becomes singular when $\gamma = \gamma_c = 1/(1+\eta)$, as (3.3.3) then has a pole at $\Omega = 0$. The approach to (3.3.3) at decreasing α is plotted in figure 3.2 (left).

Case $\eta = 1$.

To explain the rescaling procedure for understanding the approach to the singularity at $\gamma = \gamma_c$, we first focus on the case $\eta = 1$. From (3.2.54) one then has an algebraic equation for the power spectrum $C(\Omega)$

$$-\Omega^2 = -\frac{1}{C} + \frac{\alpha}{1+C} + \frac{(\gamma p)^2}{1+(\gamma p)^2 C} + \frac{p^2}{[1+2(\gamma p)^2 C]^2} \tag{3.3.4}$$

The distance from the singularity is controlled by α itself and $\delta\gamma = \gamma_c - \gamma$. Approaching the singularity along one of these two directions we get two distinct power law divergences of $C(0)$; a third direction of approach is from nonzero Ω at $\alpha = \delta\gamma = 0$.

Approaching along the $\delta\gamma$-direction ($\delta\gamma \to 0$ at $\alpha = \Omega = 0$) we find from (3.3.4)

$$C(0) \sim \frac{1}{p^2\sqrt{\delta\gamma}} \tag{3.3.5}$$

For $\alpha \to 0$ at $\Omega = 0$ and $\gamma = \gamma_c$, with $\gamma_c = 1/2$ as $\eta = 1$, the result is

$$C(0) \sim \frac{\alpha^{-\frac{1}{3}}}{p^2} \tag{3.3.6}$$

Finally for $\Omega \to 0$ at $\alpha = \delta\gamma = 0$ the low frequency tail of $C(\Omega)$ is known from chapter 2 as

$$C(\Omega) \sim \frac{1}{\sqrt{\Omega}} \tag{3.3.7}$$

These three power laws can be combined into the scaling

$$C(\Omega) = \frac{1}{p^2\sqrt{\delta\gamma}}\,\bar{C}(\bar{\Omega}, \bar{\alpha}) \tag{3.3.8}$$

$$\bar{\Omega} = \frac{\Omega}{2\,p\,\delta\gamma} \tag{3.3.9}$$

$$\bar{\alpha} = \frac{\alpha}{16\,\delta\gamma^{\frac{3}{2}}} \tag{3.3.10}$$

where the exponents of $\delta\gamma$ in $\bar{\Omega}$ and $\bar{\alpha}$ are fixed by the following standard arguments. We choose as scaling quantity $\delta\gamma$ and thus we need to understand how α scales with it. The joint dependence of $C(0)$ on α and $\delta\gamma$ becomes a dependence only on the ratio $\bar{\alpha} \sim \frac{\alpha}{\delta\gamma^{\Delta}}$ and expressed by $\bar{C}(0, \bar{\alpha})$

$$C(0) = \frac{1}{p^2\sqrt{\delta\gamma}}\,\bar{C}\Big(0, \frac{\alpha}{\delta\gamma^{\Delta}}\Big) \tag{3.3.11}$$

We postulate a power law relation defined by an exponent Δ, whose numerical value can be set by looking at the right asymptotics of $\bar{C}(0, \bar{\alpha})$ as a function of α. In fact

$$\bar{C}(0, 0) = \text{const} \tag{3.3.12}$$

$$\bar{C}(0, \alpha)|_{\alpha\to\infty} \sim \frac{\alpha^{-\frac{1}{3}}}{p^2} \tag{3.3.13}$$

In other words, at $\alpha = 0$ one should retrieve (3.3.5), and for $\alpha \to \infty$ one should recover the $\gamma = \gamma_c$ case (3.3.6). These requirements enable one to fix Δ as for $\alpha \to \infty$ one has

$$C(0) \sim \frac{1}{p^2\sqrt{\delta\gamma}}\frac{\alpha^{-\frac{1}{3}}}{\delta\gamma^{\Delta(-\frac{1}{3})}} \equiv \frac{\alpha^{-\frac{1}{3}}}{p^2} \Leftrightarrow \Delta = \frac{3}{2} \tag{3.3.14}$$

thus

$$\alpha \sim \delta\gamma^{\frac{3}{2}} \tag{3.3.15}$$

from which a renormalized $\bar{\alpha}$ can be derived. We implement in (3.3.4) the rescalings dependent on $\delta\gamma$. Inserting the ansatz (3.3.8), (3.3.9) and (3.3.10) into (3.3.4) and taking the limit $\delta\gamma \to 0$ gives

$$-1 + \bar{C}^2 + \bar{\alpha}\,\bar{C}^3 + \frac{\bar{\Omega}^2}{4}\,\bar{C}^4 = 0 \tag{3.3.16}$$

This equation implicitly determines the master curve, i.e. the scaling function $\bar{C}(\bar{\Omega}, \bar{\alpha})$. This master curve describes the power spectrum in the region where $\delta\gamma$, α and Ω are all small but $\bar{\Omega}$ and $\bar{\alpha}$ are finite, which requires in particular that the dimensionless frequency Ω must be of the order of $\delta\gamma$. Consistently with (3.3.6), one verifies that

$$\bar{C}|_{\bar{\alpha}\to\infty} \sim \bar{\alpha}^{-\frac{1}{3}} \tag{3.3.17}$$

as (3.3.16) reduces to $-1 + \bar{\alpha}\,\bar{C}^3 = 0$. At $\bar{\alpha} = 0$, one obtains

$$\bar{C}(\bar{\Omega}, 0) = \sqrt{\frac{2}{1 + \sqrt{1 + \bar{\Omega}^2}}} \tag{3.3.18}$$

which is the master curve given by implementing the same rescalings on (3.3.3) with $\eta = 1$, as expected.

Case $-1 < \eta < 1$.

Let us now consider $\eta < 1$, for which we need to work with the entire system (3.3.2). For $\delta\gamma \to 0$ at $\alpha = 0$ one finds amplitudes $C(0) \sim (1-\eta)/2\,\delta\gamma\, p^2$ and $\mathcal{B}(0) \sim -4\,\delta\gamma^2(1+\eta)^4/p^4$. At $\delta\gamma = \Omega = 0$, we get $C(0) \sim \alpha^{-1/2}$. For the third direction we can read off from chapter 2 that at $\alpha = \delta\gamma = 0$, the low-frequency tail of the power spectrum is given by $1/\Omega^2$ for $\eta < 1$. Comparing the first and third expression suggests to define a crossover frequency from $1/\Omega^2 = (1-\eta)/2\,\delta\gamma\, p^2$, giving $\Omega \sim p\sqrt{2\,\delta\gamma}$. We can fix also the exponent expressing the dependence of α on $\delta\gamma$; apart from constants, in analogy to previous explanations one has

$$C(0) \sim \frac{1}{\delta\gamma}\bar{C}\Big(0, \frac{\alpha}{\delta\gamma^{\Delta}}\Big) \tag{3.3.19}$$

with $\Delta = 2$ to obtain $\bar{C}(0,\alpha)|_{\alpha\to\infty} \sim \alpha^{-\frac{1}{2}}$, thus $\alpha \sim \delta\gamma^2$. Using this we define scaling functions for C and $\mathcal{B}$ as

$$C(\Omega) = \frac{1-\eta}{2\,\delta\gamma\, p^2}\bar{C}(\bar{\Omega},\bar{\alpha}) \tag{3.3.20}$$

$$\mathcal{B}(\Omega) = -\frac{4\,\delta\gamma^2(1+\eta)^4}{1-\eta^2}\bar{\mathcal{B}}(\bar{\Omega},\bar{\alpha}) \tag{3.3.21}$$

$$\bar{\Omega} = \frac{\Omega}{\Omega^*} = \frac{\Omega}{p\,(1-\eta)}\sqrt{\frac{1+\eta}{2\,\delta\gamma}} \tag{3.3.22}$$

$$\bar{\alpha} = \frac{\alpha\,(1-\eta)}{4\,\delta\gamma^2(1+\eta)^3} \tag{3.3.23}$$

The somewhat complicated looking prefactors are chosen here to give scaling functions that will be independent of p and η. The response $\mathcal{R}$, which also features in the original equations (3.3.2), does not need to be rescaled as it turns out to be equal to unity to leading order.

We insert the above rescalings into the system (3.3.2) and again look at the limit $\delta\gamma \to 0$. Some care is needed as there are competing orders of $\delta\gamma$ in the equations so that one has to expand the response $\mathcal{R}$ not just to $O(1)$ but to $O(\delta\gamma)$. One then finds simply $\bar{\mathcal{B}}(\bar{\Omega},\bar{\alpha}) = \bar{\alpha}$ and this makes sense: at $\alpha = 0$ we must retrieve the results of chapter 2, where the normalization of the MSRJD path integral leads all moments of auxiliary variables to vanish. The master curve for the posterior covariance spectrum can also be obtained explicitly, as

$$\bar{C}(\bar{\Omega},\bar{\alpha}) = \frac{1}{2\bar{\alpha}}\Big[-(1+\bar{\Omega}^2) + \sqrt{4\bar{\alpha}+(1+\bar{\Omega}^2)^2}\Big] \tag{3.3.24}$$

It has the limits $\bar{C}|_{\bar{\alpha}\to 0} \sim 1/(1+\bar{\Omega}^2)$, $\bar{C}|_{\bar{\Omega}\to\infty} \sim 1/\bar{\Omega}^2$ and $\bar{C}|_{\bar{\alpha}\to\infty} \sim 1/\sqrt{\bar{\alpha}}$. The latter tells us how the prediction error decreases in the regime where α is still small but larger than $\delta\gamma^2$. The fit provided by the master curve (3.3.24) for different small values of α is shown in figure 3.2 (left).

Crossover at $\eta \approx 1$.

Above we found different power law behaviours and scaling functions for $\eta = 1$ and $\eta < 1$, thus a crossover must occur at $\eta \approx 1$. To see this, the two cases $\eta = 1$ and $\eta < 1$ can be analyzed as limit cases of a more general scaling ansatz that accounts explicitly for the effect of $\epsilon = 1 - \eta$, i.e. the distance from the symmetric value $\eta = 1$. This becomes an additional critical parameter that enters the scaling functions in the combination

$$\bar{\epsilon} = \epsilon / \sqrt{\delta\gamma}$$

We define these scaling functions via

$$C(\Omega) = \frac{(2+\bar{\epsilon})}{2\sqrt{\delta\gamma}\, p^2} \bar{C}(\bar{\Omega}, \bar{\alpha}, \bar{\epsilon}) \tag{3.3.25}$$

$$\mathcal{B}(\Omega) = -\frac{32\,\delta\gamma^{\frac{3}{2}}}{2+\bar{\epsilon}} \bar{\mathcal{B}}(\bar{\Omega}, \bar{\alpha}, \bar{\epsilon}) \tag{3.3.26}$$

$$\bar{\Omega} = \frac{\Omega}{p\,(2+\bar{\epsilon})\delta\gamma} \tag{3.3.27}$$

$$\bar{\alpha} = \frac{\alpha\,(2+\bar{\epsilon})}{32\,\delta\gamma^{\frac{3}{2}}} \tag{3.3.28}$$

We then find again $\bar{\mathcal{B}} = \bar{\alpha}$ while the master curve $\bar{C}(\bar{\Omega}, \bar{\alpha}, \bar{\epsilon})$ solves a fourth-order equation

$$(8 + \bar{C}\,\bar{\epsilon}(2+\bar{\epsilon}))^2(-4 + \bar{C}(2+\bar{\epsilon})(-\bar{\epsilon} + \bar{C}(1+\bar{\alpha}\,\bar{C})(2+\bar{\epsilon}))) + \bar{C}^4(2+\bar{\epsilon})^6\bar{\Omega}^2 = 0 \tag{3.3.29}$$

The solution of this for a range of different $\bar{\epsilon}$ is plotted in figure 3.3 (left).

The two previous cases $\eta = 1$ and $\eta < 1$ (with $\delta\gamma \to 0$) are recovered as the limits respectively for $\bar{\epsilon} \to 0$ and $\bar{\epsilon} \to \infty$. In the first limit, $(2+\bar{\epsilon})\sqrt{\delta\gamma} \sim 2\sqrt{\delta\gamma}$ and the rescaling relations (3.3.8), (3.3.9), (3.3.10) for $\eta = 1$ are retrieved as they should be; accordingly, the equation (3.3.29) for the master curve becomes exactly (3.3.16). On the other hand, when $\bar{\epsilon} \to \infty$, $(2+\bar{\epsilon})\sqrt{\delta\gamma} \sim \bar{\epsilon}\sqrt{\delta\gamma} = \epsilon$ and we recover the rescalings (3.3.20)-(3.3.23) adopted for $\eta < 1$; in this case (3.3.29) reduces to

$$-1 + \bar{C} + \bar{\alpha}\bar{C}^2 + \bar{C}\bar{\Omega}^2 = 0 \tag{3.3.30}$$

whose positive solution is given by (3.3.24).

3.3.4 Master curves for $\alpha \to 1$ and $p \to 0$

In this section we look at the scaling around the second critical region, $\alpha \to 1$ (i.e. $N^{\mathrm{s}} = N^{\mathrm{b}}$) and $p \to 0$ (i.e. $k \gg \lambda$ at fixed $\sigma_{\mathrm{s}}/\sigma_{\mathrm{b}}$). As will be shown, the results here are independent of the degree of symmetry η of the interactions among the hidden variables. This is consistent with the intuition

that the critical behaviour is dominated by whether a direction in the hidden trajectory space is constrained by observations or not, i.e. by hidden-to-observed rather than hidden-to-hidden interactions. As previously done, we first explore the two directions of divergence separately. At $\alpha = 1$, we find the following power law scaling with p of the amplitudes

$$C(0) \sim \frac{1}{p\sqrt{\gamma^2+1}} \tag{3.3.31}$$

$$\mathcal{R}(0) \sim \frac{\gamma p}{\sqrt{\gamma^2+1}} \tag{3.3.32}$$

$$\mathcal{B}(0) \sim -1 \tag{3.3.33}$$

Approaching from the other direction, $p = 0$, gives $C(0) \sim 1/\delta\alpha$, with $\delta\alpha = \alpha - 1$. At $\alpha = 1$ and $p = 0$, finally, one finds for small nonzero frequency Ω that $C(\Omega) \sim 1/\Omega$. Equating the three divergences above identifies a crossover frequency $\Omega^* = p\sqrt{\gamma^2+1}$ and similarly a characteristic value for $\delta\alpha$: as before, we start from writing

$$C(0) \sim \frac{1}{p\sqrt{\gamma^2+1}} \bar{C}\left(0, \frac{\delta\alpha}{(p\sqrt{\gamma^2+1})^{\Delta}}\right) \tag{3.3.34}$$

One retrieves $C(0) \sim 1/\delta\alpha$ in the $p \to 0$ limit if $\Delta = 1$. We thus define rescaled quantities again

$$C(\Omega) = \frac{1}{p\sqrt{\gamma^2+1}} \bar{C}(\bar{\Omega}, \delta\bar{\alpha}) \tag{3.3.35}$$

$$\mathcal{R}(\Omega) = \frac{\gamma p}{\sqrt{\gamma^2+1}} \bar{\mathcal{R}}(\bar{\Omega}, \delta\bar{\alpha}) \tag{3.3.36}$$

$$\mathcal{B}(\Omega) = -\bar{\mathcal{B}}(\bar{\Omega}, \delta\bar{\alpha}) \tag{3.3.37}$$

$$\bar{\Omega} = \frac{\Omega}{\Omega^*} = \frac{\Omega}{p\sqrt{\gamma^2+1}} \tag{3.3.38}$$

$$\delta\bar{\alpha} = \frac{\delta\alpha}{p\sqrt{\gamma^2+1}} \tag{3.3.39}$$

Inserting into the system (3.3.2), taking $p \to 0$ and keeping only the leading terms one finds as the master curve for $C(\Omega)$

$$\bar{C}(\bar{\Omega}, \delta\bar{\alpha}) = \frac{-\delta\bar{\alpha} + \sqrt{4 + \delta\bar{\alpha}^2 + 4\bar{\Omega}^2}}{2(1+\bar{\Omega}^2)} \tag{3.3.40}$$

with limits $\bar{C}|_{\delta\bar{\alpha}\to 0} \sim 1/\sqrt{1+\bar{\Omega}^2}$, $\bar{C}|_{\bar{\Omega}\to\infty} \sim 1/\bar{\Omega}$ and $\bar{C}|_{\delta\bar{\alpha}\to\infty} \sim 1/\delta\bar{\alpha}$. From the latter one sees again the decrease of the inference error for increasing number of observations (while remaining in the regime studied here where $\delta\alpha$ is small, $p < \delta\alpha \ll 1$). The master curve predictions for generic $\delta\bar{\alpha}$ are compared to direct numerical evaluation of C in figure 3.2 (right).

So far we had focussed on the $\alpha \to 1$ end of the second critical region. As this region covers the entire range $0 < \alpha < 1$, however, one can also study the critical behaviour as $p \to 0$ for fixed $\alpha < 1$. The crossover into this region can be seen by taking $\delta\bar{\alpha} \to -\infty$, where $\bar{C} \to |\delta\bar{\alpha}|/(1+\bar{\Omega}^2)$

from (3.3.40). Including the prefactor from (3.3.35) gives

$$C(0) = \frac{1-\alpha}{p^2(\gamma^2+1)(1+\bar{\Omega}^2)} \tag{3.3.41}$$

This suggests that in general, for finite $1-\alpha$, C will be $\sim 1/p^2$. Generalizing this one finds that the appropriate scaling for small p of all order parameters is

$$C(\Omega) = \frac{1}{p^2\sqrt{\gamma^2+1}}\bar{C}(\bar{\Omega},\delta\bar{\alpha}) \tag{3.3.42}$$

$$\mathcal{R}(\Omega) = \frac{\gamma}{\sqrt{\gamma^2+1}}\bar{\mathcal{R}}(\bar{\Omega},\delta\bar{\alpha}) \tag{3.3.43}$$

$$\mathcal{B}(\Omega) = -\bar{\mathcal{B}}(\bar{\Omega},\delta\bar{\alpha}) \tag{3.3.44}$$

$$\bar{\Omega} = \frac{\Omega}{p\sqrt{\gamma^2+1}} \tag{3.3.45}$$

$$\delta\bar{\alpha} = \frac{|\delta\alpha|}{\sqrt{\gamma^2+1}} = \frac{1-\alpha}{\sqrt{\gamma^2+1}} \tag{3.3.46}$$

By substitution into the system (3.3.2) and taking the limit $p \to 0$ one can then obtain explicit solutions for $\bar{C}(\bar{\Omega},\delta\bar{\alpha})$, $\bar{\mathcal{R}}(\bar{\Omega},\delta\bar{\alpha})$, $\bar{\mathcal{B}}(\bar{\Omega},\delta\bar{\alpha})$; as before these are independent of η. We do not state the full expressions here but note that in the limit where $1-\alpha \ll 1$ one retrieves the result (3.3.41) as required for consistency between the two scaling limits.

Finally it is useful to look at the α-dependence of $C(0)$ and $C(\Omega)$ in the non-critical range $\alpha > 1$. For large α we find $C(0)|_{\alpha\to\infty} \sim 1/\alpha$: this means that with many observations the predictions for the hidden dynamics will become arbitrarily precise, in line with intuition. The decrease in $C(0)$ with increasing α can be seen in figure 3.4, where one observes also that at $\omega \sim O(1)$ all spectra collapse into a Lorentzian tail. This indicates an exponential decay of the correlations between prediction errors in the temporal domain. As the amplitude of the tail is largely α-independent, the typical time of this exponential decay decreases with α: with many observations, errors in the prediction of the hidden states become progressively less correlated with each other. In the next section we look more systematically at the information one can extract on relaxation times from the power spectra.

3.4 Posterior variance in the time domain

3.4.1 Relaxation time for $\gamma \to \gamma_c$ and $\alpha \to 0$

We look at the relaxation time, which is a measure of time correlations in the errors of inferred hidden values. We study in particular how it depends on the number of observations and the interaction parameters.

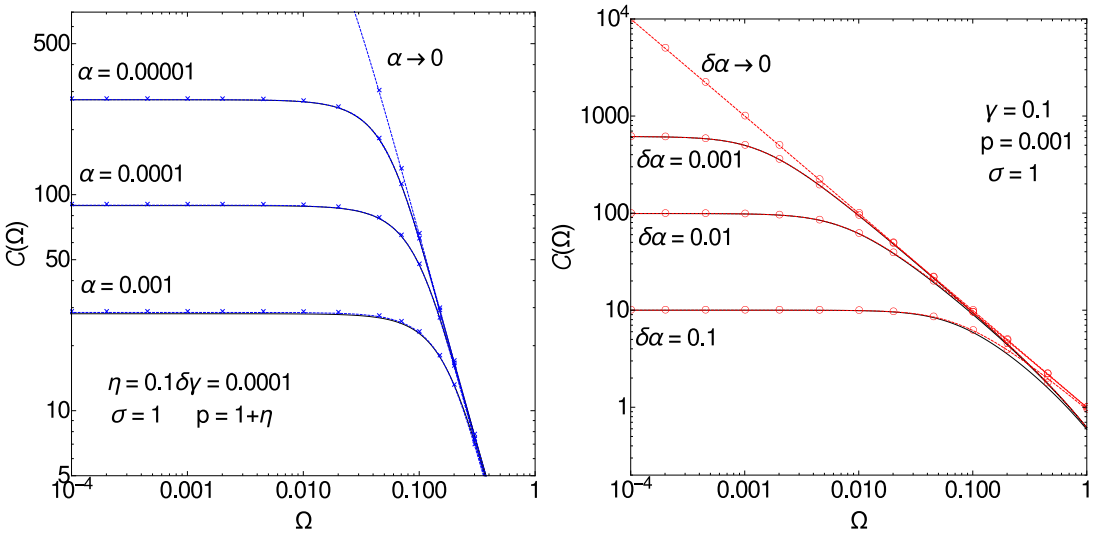

Figure 3.2: (Left) Numerical solutions for small α: the approach to the limit curve for $\alpha \to 0$ with γ close to γ_c is plotted. This master curve and the ones for $\alpha = 0.00001$, $\alpha = 0.0001$ and $\alpha = 0.001$ are given by blue dashed lines with crosses. In this way one can see the amplitude variation for relatively large $\bar{\alpha}$, whose values are of order $\sim 10^2$ (for $\alpha = 0.00001$), 10^3 (for $\alpha = 0.0001$) and 10^4 (for $\alpha = 0.001$). Even for smaller $\bar{\alpha}$, the variation in shape with this parameter is small: $\bar{\alpha}$ mainly affects the height of the plateau close to $\Omega = 0$ and the position of the crossover to the large frequency Lorentzian tail. (Right) Numerical solutions for small $\delta\alpha = \alpha - 1$: the approach to the limit curve for $\delta\alpha \to 0$ with p close to 0 is plotted. This master curve and the ones for $\delta\alpha = 0.001$, $\delta\alpha = 0.01$ and $\delta\alpha = 0.1$ are shown as red dashed lines with circles. From this plot one can examine the variation with $\delta\bar{\alpha}$, whose corresponding values are $\delta\bar{\alpha} \sim 1$ (for $\delta\alpha = 0.001$), $\delta\bar{\alpha} \sim 10$ (for $\delta\alpha = 0.01$) and $\delta\bar{\alpha} \sim 100$ (for $\delta\alpha = 0.1$).

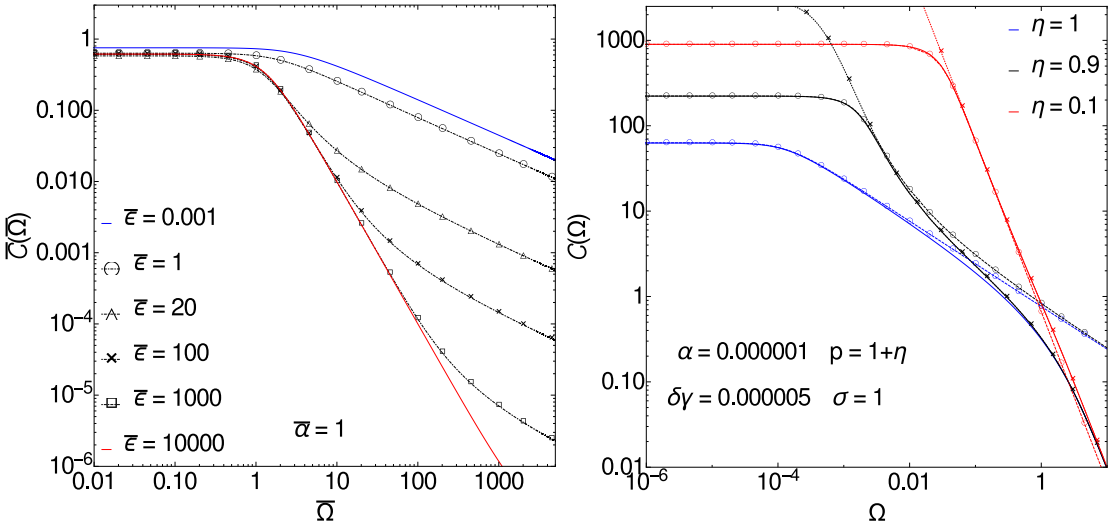

Figure 3.3: (Left) Master curves for different values of $\bar{\epsilon}$, the parameter indicating effectively the distance from $\eta = 1$ when $\delta\gamma$ is close to zero. One can see the $1/\sqrt{\bar{\Omega}}$ tail for $\bar{\epsilon} \to 0$ and the $1/\bar{\Omega}^2$ one for $\bar{\epsilon} \to \infty$ (blue and red curves), while intermediate values of $\bar{\epsilon}$ show a crossover between these two tails. (Right) Numerics for different values of η, the symmetry parameter, at $\alpha \to 0$ and $\delta\gamma \to 0$. For $\eta = 0.1$, $\Omega^* \sim 3 \cdot 10^{-3}$ and for $\Omega^* \ll \Omega \ll 1$ one sees the Lorentzian tail. For $\eta = 1$, $\Omega^* \sim 5 \cdot 10^{-6}$ and in the range $\Omega^* \ll \Omega \ll 1$ one has the $\sim 1/\sqrt{\Omega}$ tail. For $\eta = 0.9$ (i.e. $\epsilon = 1 - \eta = 0.1$, $\bar{\epsilon} \approx 45$), the results interpolate between these two regimes as expected from the left figure; the crossover occurs at $\Omega \sim 0.01$. In fact, the limit $\bar{C}|_{\bar{\Omega}\to\infty} \sim 1/\bar{\Omega}^2$, once one reinserts the dependence on $1 - \eta = \epsilon$, gives $C(\Omega) \sim \epsilon^3/\Omega^2$ for $\Omega^* \ll \Omega \ll 1$, while (3.3.16) has a tail $C(\Omega) \sim 2/\sqrt{\Omega}$: these two tails meet around $\Omega \sim \epsilon^2$ ($\epsilon^2 = 0.01$ in this case), in agreement with the results for the limit $\epsilon \ll 1$ in the case without observations, i.e. chapter 2. The dashed lines with circles are master curves for Ω in the vicinity of Ω^* for the different η. Dotted lines with crosses trace the master curves at $\Omega \sim O(1)$, which are independent of α and equal to the curves at $\alpha = 0$. In the relevant range of large frequencies they behave essentially as Lorentzians.

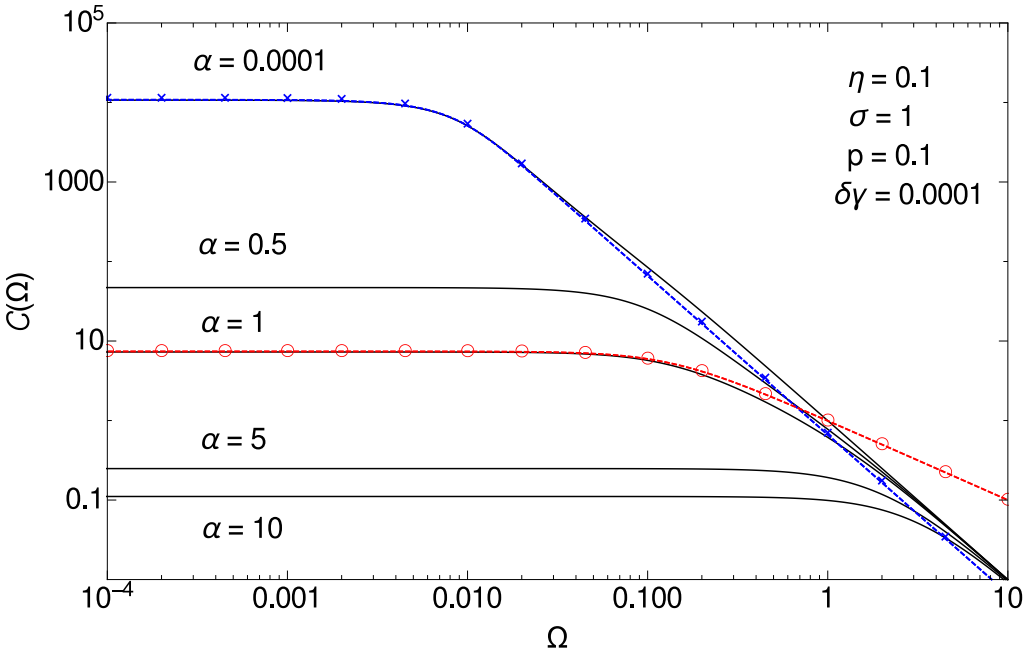

Figure 3.4: $C(\Omega)$ for different α, at small fixed η. We have chosen γ and p close to their critical values $1+\eta$ and 0, respectively, in order to see both critical regions as α is increased. The master curves for $\alpha \to 0$ and $\alpha \to 1$, resulting from the critical rescalings, are plotted (blue dashed line with crosses and red dashed line with circles respectively): $\Omega^* \sim 10^{-3}$ for the $\alpha \to 0$ master curve, while $\Omega^* \sim 10^{-1}$ for the one for $\alpha \to 1$. For $\Omega^* \ll \Omega \ll 1$ one has a Lorentzian tail $C \sim 1/\Omega^2$ for α small and α big, while a different power-law feature, namely $C \sim 1/\Omega$, emerges for $\alpha \to 1$. At $\Omega \sim O(1)$ a crossover to Lorentzian behaviour is seen for any α.

The relaxation time can be defined in a mean-squared sense as

$$\tau^2 = \frac{\int_{-\infty}^{+\infty} t^2 C(t) dt}{2\,\tilde{C}(0)} = -\frac{1}{2\,C(0)} \frac{d^2 C(\omega)}{d^2\omega}\bigg|_{\omega=0} = -\frac{1}{2\,\bar{C}(0,\bar{\alpha})} \frac{d^2 \bar{C}(\bar{\Omega},\bar{\alpha})}{d^2\omega}\bigg|_{\omega=0} \tag{3.4.1}$$

As for the other quantities, let us introduce the dimensionless version of this typical timescale

$$\mathcal{T} = \sigma\tau \tag{3.4.2}$$

and look separately at the two critical regions, starting in this subsection with the first one. As τ is determined directly from the power spectrum, its scaling form follows from that of $C(\Omega)$.

Case $\eta = 1$.

The relaxation time is rescaled as

$$\mathcal{T} = \mathcal{T}^*\,\bar{\tau}(\bar{\alpha}) = \frac{1}{p\,\delta\gamma}\,\bar{\tau}(\bar{\alpha}) \tag{3.4.3}$$

where, by the definition itself in terms of frequency (3.4.1), one has $\mathcal{T}^* \sim 1/\Omega^*$ and $\bar{\tau}(\bar{\alpha})$ is the solution of a system of two equations

$$\begin{cases} (2 + 3\,\bar{\alpha}\,\bar{C})\,\bar{C}\,\bar{\tau}^2 - \frac{1}{4}\bar{C}^3 = 0 \\ -1 + \bar{C}^2 + \bar{\alpha}\,\bar{C}^3 = 0 \end{cases} \tag{3.4.4}$$

which can be obtained by deriving from (3.3.16) one equation for $\bar{C}(0,\bar{\alpha})$ and one for $d^2\bar{C}(\bar{\Omega},\bar{\alpha})/d^2\omega\big|_{\omega=0}$ and using relations (3.4.1) and (3.4.3). For simplicity we have denoted $\bar{C}(0,\bar{\alpha})$ and $\bar{\tau}(\bar{\alpha})$ as $\bar{C}$ and $\bar{\tau}$. From these two equations we note $\bar{C}(0,\bar{\alpha})|_{\bar{\alpha}\to\infty} \sim \bar{\alpha}^{-\frac{1}{3}}$ and $\bar{\tau}|_{\bar{\alpha}\to\infty} \sim \bar{\alpha}^{-\frac{2}{3}}$, which implies for $\delta\gamma^2 < \alpha \ll 1$

$$\mathcal{T} \sim \alpha^{-\frac{2}{3}} \tag{3.4.5}$$

This power law dependence is visible in figure 3.5 (right, see $\eta = 1$ curve).

Case $-1 < \eta < 1$.

By applying (3.4.1) with (3.3.24), one can rescale in this regime according to

$$\mathcal{T} = \frac{1}{(1-\eta)\,p}\sqrt{\frac{1+\eta}{2\,\delta\gamma}}\,\bar{\tau}(\bar{\alpha}) \tag{3.4.6}$$

and $\bar{\tau}(\bar{\alpha})$ is given by

$$\bar{\tau}(\bar{\alpha}) = \frac{1}{(4\bar{\alpha}+1)^{\frac{1}{4}}} \tag{3.4.7}$$

with limit $\bar{\tau}|_{\bar{\alpha}\to\infty} \sim \bar{\alpha}^{-\frac{1}{4}}$ corresponding to

$$\mathcal{T} \sim \alpha^{-\frac{1}{4}} \qquad \delta\gamma^2 < \alpha \ll 1 \tag{3.4.8}$$

This dependence can be verified in figure 3.5 (right, see curve for $\eta = 0.1$).

Crossover at $\eta \approx 1$.

The relaxation time scalings above can be seen as limit cases of a more general scaling linked to the parameter $\bar{\epsilon}$. From the general master curve (3.3.29) we can derive the following system of two equations

$$\begin{cases} [8 + \bar{C}\bar{\epsilon}(2+\bar{\epsilon})]^2[-4 + \bar{C}(2+\bar{\epsilon})(-\bar{\epsilon} + \bar{C}(1+\bar{\alpha}\bar{C})(2+\bar{\epsilon}))] = 0 \\ [-16\bar{\epsilon} + \bar{C}(2+\bar{\epsilon})(16 - 3\bar{\epsilon}^2 + 4\bar{C}\bar{\epsilon}(2+\bar{\epsilon}) + \bar{\alpha}\bar{C}(24 + 5\bar{C}\bar{\epsilon}(2+\bar{\epsilon}))]\bar{\tau}^2 = \frac{\bar{C}^3(2+\bar{\epsilon})\bar{\epsilon}^4}{(8+\bar{C}\bar{\epsilon}(2+\bar{\epsilon}))} \end{cases} \tag{3.4.9}$$

where $\bar{C}$ is shorthand for $\bar{C}(0, \bar{\alpha}, \bar{\epsilon})$ and $\bar{\tau}$ for $\bar{\tau}(\bar{\alpha}, \bar{\epsilon})$. Consistent with the role of $\bar{\epsilon}$ discussed above, one can check that the limit for $\bar{\epsilon} \to 0$ is precisely the system (3.4.4), while for $\bar{\epsilon} \to \infty$ (3.4.9) becomes

$$\begin{cases} -1 + \bar{C} + \bar{\alpha}\bar{C}^2 = 0 \\ -\bar{C} + \bar{\tau}^2(-3 + \bar{C}(4 + 5\bar{\alpha}\bar{C})) = 0 \end{cases} \tag{3.4.10}$$

Here the first equation agrees with (3.3.24) as it should. Solving for $\bar{\tau}$, one finds

$$\bar{\tau}(\bar{\alpha}) = \frac{1}{(1+4\bar{\alpha})^{\frac{1}{4}}} \tag{3.4.11}$$

For generic $\bar{\epsilon}$, the rescaled relaxation time $\bar{\tau}(\bar{\alpha})$ must then exhibit a crossover in its large $\bar{\alpha}$ power law behaviour, i.e. from $\bar{\alpha}^{-\frac{1}{4}}$ to $\bar{\alpha}^{-\frac{2}{3}}$. This crossover takes place around $\bar{\alpha}^* = \bar{\epsilon}^2/4$; see figure 3.5 (left).

3.4.2 Relaxation time for $\alpha \to 1$ and $p \to 0$

We discuss briefly the behaviour of the relaxation time in the second critical region. From the rescaling of C one has

$$\mathcal{T} = \frac{1}{p\sqrt{(1+\gamma^2)}}\,\bar{\tau}(\delta\bar{\alpha}) \tag{3.4.12}$$

and

$$\bar{\tau}(\delta\bar{\alpha}) = \sqrt{\frac{1}{2}\Big(1 - \frac{\delta\bar{\alpha}}{\sqrt{4+\delta\bar{\alpha}^2}}\Big)} \tag{3.4.13}$$

with limit $\bar{\tau}|_{\delta\bar{\alpha}\to\infty} \sim 1/\delta\bar{\alpha}$ giving

$$\mathcal{T} \sim \frac{1}{\delta\alpha} \qquad p < \delta\alpha \ll 1 \tag{3.4.14}$$

The $\delta\alpha$ dependence of $\mathcal{T}$ is shown in figure 3.6 (left). As a general unifying feature of the relaxation times, one can see that they decrease significantly with increasing α (figure 3.6, right): as the values of the hidden variables become constrained increasingly strongly by those of the observed ones, the remaining uncertainty in the prediction becomes local in time.

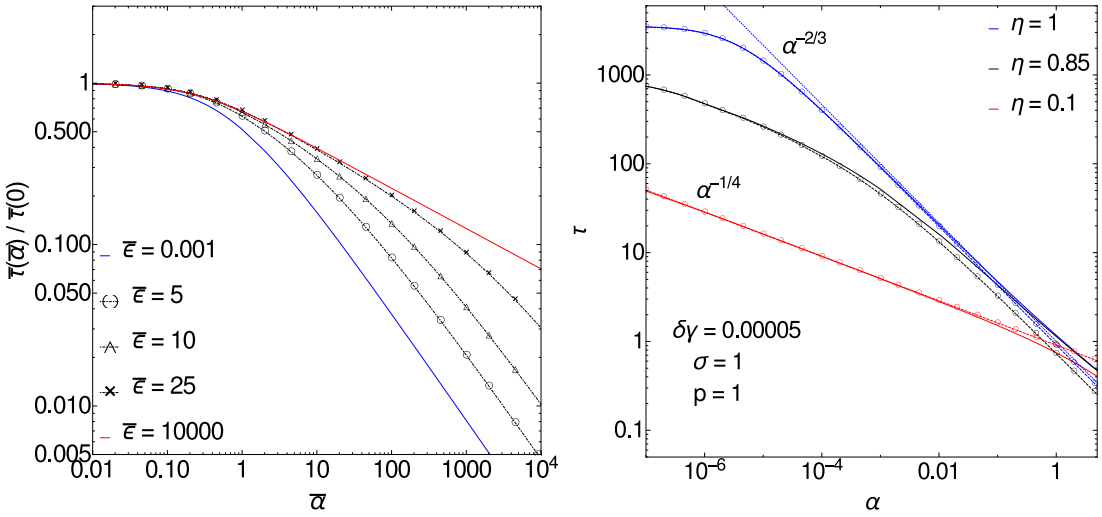

Figure 3.5: (Left) The rescaled relaxation time $\bar{\tau}(\bar{\alpha})$, normalized at the origin, for $\alpha \to 0$ and $\gamma \to \gamma_c$ as a function of $\bar{\alpha}$ for different values of $\bar{\epsilon}$. The two limits, $\bar{\epsilon} \to \infty$ and $\bar{\epsilon} \to 0$, are highlighted in red and blue respectively. For intermediate values of $\bar{\epsilon}$ one can see an interpolation between the corresponding power laws, i.e. $\bar{\alpha}^{-\frac{1}{4}}$ and $\bar{\alpha}^{-\frac{2}{3}}$, and this crossover occurs at $\bar{\alpha}^* = \bar{\epsilon}^2/4$ (for instance, at $\bar{\epsilon} = 25$, $\bar{\alpha}^* \sim 150$). (Right) Relaxation time in the vicinity of the critical region $\alpha \to 0$ and $\gamma \to \gamma_c$, as a function of α and for different η. Solid lines are the numerics, dashed ones with circles the analytic master curves. A plateau for small α emerges for η close to 1. For $\delta\gamma^2 < \alpha \ll 1$ one can see $\tau \sim \alpha^{-\frac{2}{3}}$ for $\eta = 1$ and $\tau \sim \alpha^{-\frac{1}{4}}$ for $\eta = 0.1$, as suggested by eqs. (3.4.5) and (3.4.8). The case $\eta = 0.85$ interpolates between these power tails with a crossover at $\alpha \sim 0.0005$; this corresponds to the crossover on the left for $\bar{\epsilon} \sim 20$.

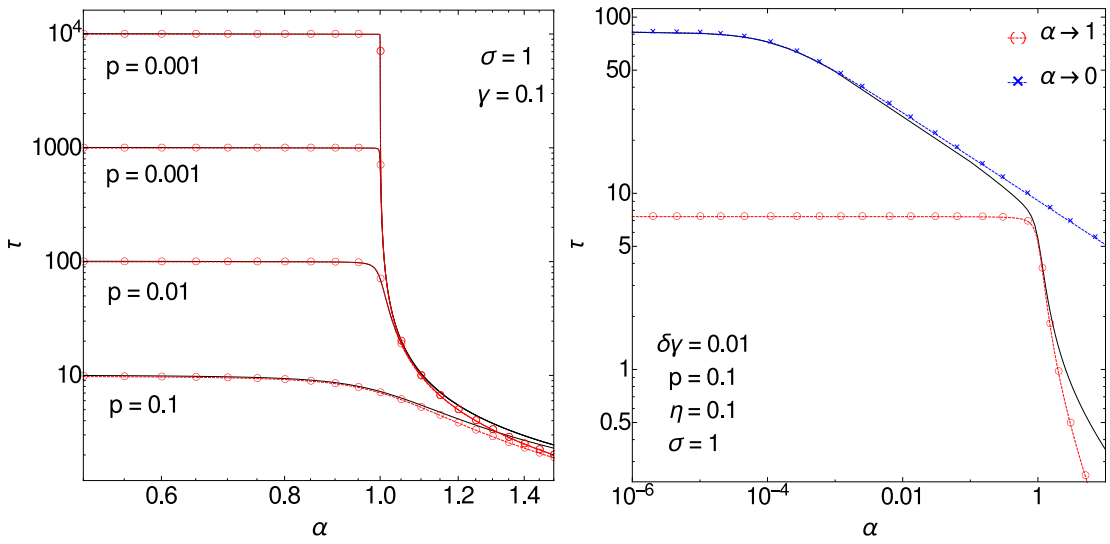

Figure 3.6: (Left) Relaxation time in the vicinity of the critical region $\alpha \to 1$ and $p \to 0$, as a function of α. As the variation with γ is weak in this regime, we fix $\gamma = 0.1$. Solid lines are the numerics, dashed ones with circles the analytic master curves. One can see that τ stays roughly constant for $\alpha < 1$ while it drops to smaller and smaller values for increasing $\alpha > 1$. (Right) Relaxation time for small p and $\delta\gamma$ as a function of α: an interpolation between the behaviours in the left plot and in figure 3.5 (right) can be seen here. We stress that the red master curve with circles is expected to give a good fit for $\alpha \approx 1$ only, consistent with our results.

3.4.3 Equal time posterior variance for $\gamma \to \gamma_c$ and $\alpha \to 0$

Finally we turn to the behaviour of the inference error for the prediction of hidden unit trajectories. This is given by the equal time posterior correlator

$$C(t-t) = C(0) = \int_{-\infty}^{\infty} \tilde{C}(\omega) d\omega = \frac{\sigma_s^2}{k^2} \sigma \int_{-\infty}^{\infty} C(\Omega) d\,\Omega = \frac{\sigma_s \sigma_b}{k} C_0 \tag{3.4.15}$$

where $C_0 = \int_{-\infty}^{\infty} C(\Omega) d\,\Omega$ is a dimensionless equal time posterior variance. We see that the size of the error is generically proportional to the noise acting on the dynamics of hidden and observed variables, and inversely proportional to the hidden-to-observed interaction strength. Its critical scaling properties depend on whether the integral over Ω that defines C_0 is dominated by small frequencies $\Omega \sim \Omega^*$, where Ω^* is the relevant frequency scale in the appropriate critical region, or by $\Omega \sim 1$. One has the first case when the critical master curve for $C(\Omega)$ has an integrable tail towards large (scaled) frequencies, otherwise the second.

Case $\eta = 1$.

Here the dominant contribution to the integral (3.4.15) comes from $\Omega \sim O(1)$, and in this region the master curve for the spectrum is given by (3.3.3) (the case without observations). As a consequence we expect the equal time correlator to become essentially independent of α for small α, as one can see in figure 3.7 (left, curves for $\eta = 1$). More generally this implies that the dependence on α is smooth, and to leading order unaffected by the vicinity of the critical region.

Case $-1 < \eta < 1$.

Here the dominant contribution to the prediction error comes from $\Omega \sim \Omega^*$, thus one has to evaluate the integral of the master curve (3.3.24). To do so we note from (3.3.20) that C_0 can be written in scaled form as

$$C_0 = \frac{(1-\eta)\Omega^*}{2\,\delta\gamma\,p^2}\,\bar{C}_0(\bar{\alpha}) \tag{3.4.16}$$

with $\bar{C}_0(\bar{\alpha}) = \int_{-\infty}^{\infty} \bar{C}(\bar{\Omega}, \bar{\alpha}) d\,\bar{\Omega}$. This function encodes the entire α-dependence in the critical region. It has a finite limit for $\bar{\alpha} \to 0$ while for large $\bar{\alpha}$ it decays as $\bar{C}_0 \sim \bar{\alpha}^{-1/4}$ as one can show by noting that the relevant frequencies $\bar{\Omega}$ in (3.3.24) are then $\sim \bar{\alpha}^{-1/4}$. Using this asymptotic behaviour in (3.4.16) and substituting also the expressions for Ω^* and $\bar{\alpha}$ from (3.3.22) and (3.3.23), respectively, one obtains

$$C(0) \sim \frac{(1-\eta)^2}{p\,\sqrt{2\delta\gamma(1+\eta)}} \bar{\alpha}^{-\frac{1}{4}} = \frac{1}{p}(1-\eta)^{\frac{7}{4}}(1+\eta)^{\frac{1}{4}}\alpha^{-\frac{1}{4}} \tag{3.4.17}$$

We thus predict $C(0) \sim \alpha^{-1/4}$, and this is consistent with the numerics, see e.g. figure 3.7 (left, curves for $\eta = 0.1$), where $\bar{\alpha} \gg 1$ corresponds to $\delta\gamma^2 \ll \alpha \ll 1$ from (3.3.23): in this range the zero frequency amplitude is then independent of $\delta\gamma$, and the value of the latter appears only as the lower limit of the range where this result applies. The power law behaviour $C(0) \sim \alpha^{-1/4}$ is also consistent with the scaling $\tilde{C}(0) \sim \tau C(0)$ that one would generally expect, barring any exceptions due to strongly non-exponential correlations: recall here that we found previously $\tau \sim \alpha^{-1/4}$ and $\tilde{C}(0) \sim \alpha^{-1/2}$.

3.4.4 Equal time posterior variance for $\alpha \to 1$ and $p \to 0$

In the second critical region and focussing on $\alpha \to 1$, we have an interesting marginal case where the equal-time variance (3.4.15) has contributions from all frequencies ranging from the critical frequency scale $\Omega \sim \Omega^*$ to $\Omega \sim 1$. This is because the power spectrum (3.3.40) for critical frequencies has a $1/\bar{\Omega}$ tail for large $\bar{\Omega} = \Omega/\Omega^*$, which gives a logarithmically divergent integral (3.4.15). This divergence is cut off only by the crossover to a Lorentzian tail when $\Omega = O(1)$. Including the prefactor from (3.3.35), one thus estimates

$$C_0 \approx \frac{\Omega^*}{p\sqrt{\gamma^2+1}} 2 \int_0^{1/\Omega^*} \bar{C}(\bar{\Omega}, \delta\bar{\alpha}) d\bar{\Omega} \tag{3.4.18}$$

The fraction in front of the integral equals unity as $\Omega^* = p\sqrt{\gamma^2+1}$ from (3.3.38) so from the $1/\bar{\Omega}$ tail one finds $C_0 \approx 2\ln(1/\Omega^*)$ to leading order. All of the interesting dependence on α is in the next subleading term, which is relevant in practice as it only competes with a logarithmic divergence. Writing $\ln(1/\Omega^*)$ as $\int_0^{1/\Omega^*} d\bar{\Omega}/(1+\bar{\Omega})$, this subleading term can be split off in the form

$$C_0 \approx 2\ln(1/\Omega^*) + 2\int_0^{\infty} [\bar{C}(\bar{\Omega}, \delta\bar{\alpha}) - (1+\bar{\Omega})^{-1}] d\bar{\Omega} \tag{3.4.19}$$

The remaining integral is convergent at the upper limit so we have taken the upper limit $1/\Omega^*$ to infinity as is appropriate to get the leading contribution for $p \to 0$. The integral is then a function of $\delta\bar{\alpha}$ only, which one finds varies as $|\delta\bar{\alpha}|$ for $\delta\bar{\alpha} \to -\infty$ and as const. $- \ln(\delta\bar{\alpha})$ for $\delta\bar{\alpha} \to \infty$. The two dominant terms (3.4.19) are plotted (red line) in figure 3.8 where the linear scale inset clearly shows the linear dependence on $|\delta\bar{\alpha}| \propto 1 - \alpha$.

As a common trend across the two critical regions we have the intuitively reasonable result that the inference error decreases when the number of observed variables gets bigger, see figure 3.7 (right).

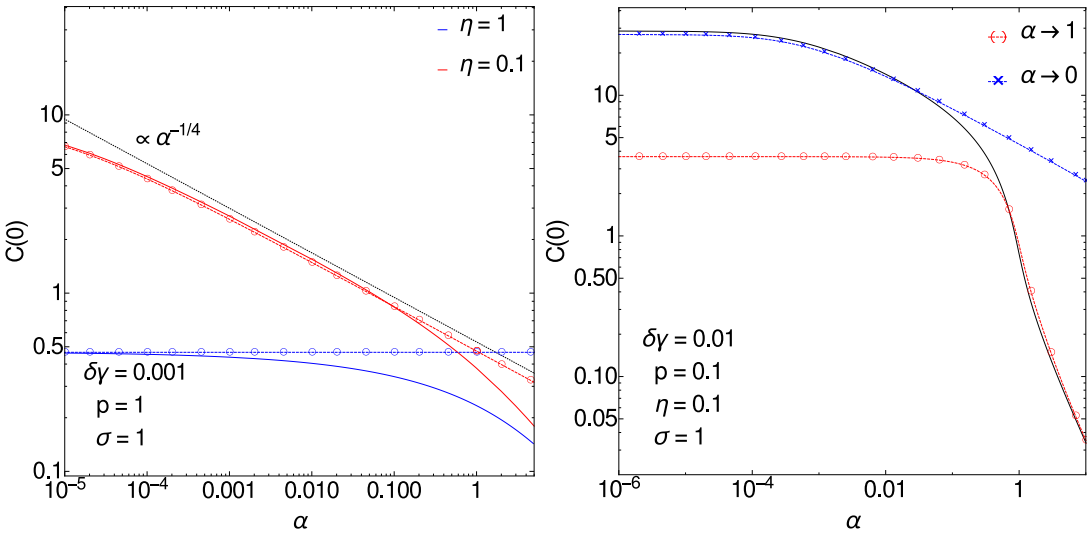

Figure 3.7: (Left) Inference error in the vicinity of the critical region $\alpha \to 0$ and $\gamma \to \gamma_c$, as a function of α and for different η. Solid lines are the numerics, dashed ones with circles the analytic master curves. The dominant contribution to the inference error is given by frequencies $\Omega \sim \Omega^*$ for $\eta = 0.1$ while it is given by $\Omega \sim O(1)$ for $\eta = 1$. The master curve in this latter regime is the α-independent posterior variance of the case without observations, i.e. a straight line, which $C(0)$ approaches for $\alpha \to 0$ as it should; for larger α it exhibits a smooth dependence on α unconnected to any critical behaviour. (Right) Inference error for small p and $\delta\gamma$ as a function of α. This connects the behaviours in the left plot and in figure 3.8. The red master curve with circles is expected to give a good fit only around $\alpha \approx 1$, as observed.

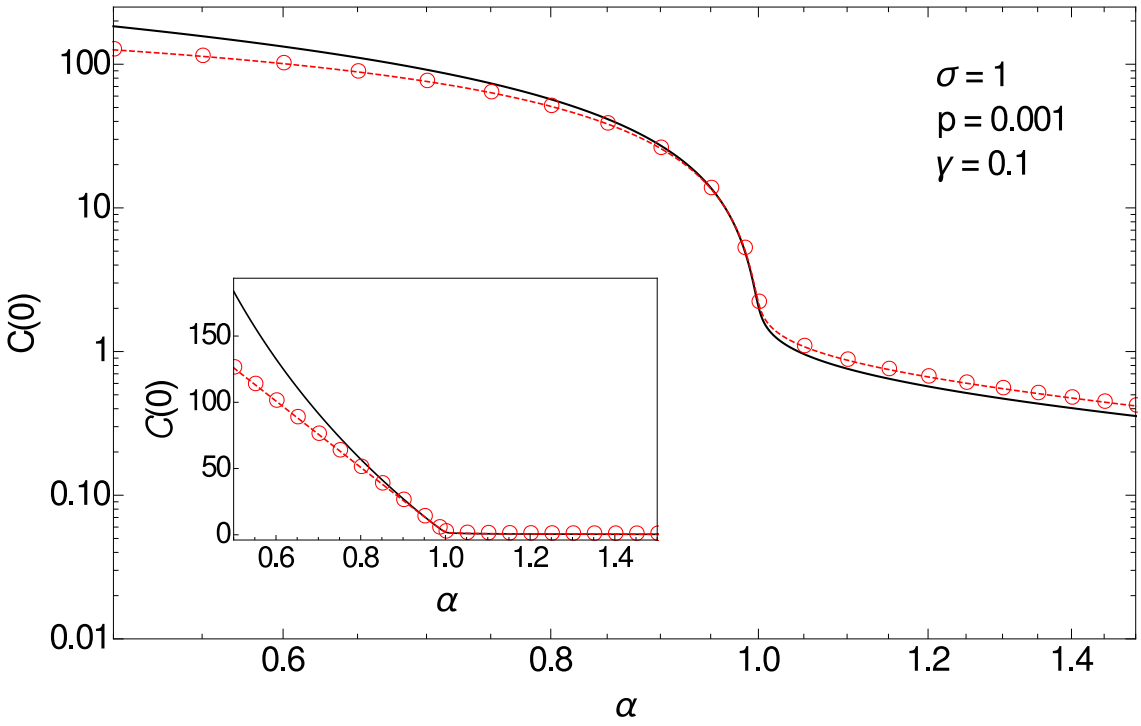

Figure 3.8: Inference error in the vicinity of the critical region $\alpha \to 1$ and $p \to 0$, as a function of α, at fixed $\gamma = 0.1$. The dashed line with circles shows the prediction from the logarithmic divergence and first subleading term (3.4.19), which is qualitatively remarkably accurate even away from $\alpha = 1$. The prediction behaves as $\sim (1 - \alpha)$ for $\alpha < 1$ as the linear scale inset shows.

3.5 Discussion and Conclusion

In this chapter we considered a network of continuous degrees of freedom where some nodes are hidden and the others observed. In such a setting we discussed an application of the Extended Plefka Expansion to the problem of inferring hidden states over time when one has seen the dynamics of the observed ones. In particular we focussed on the case of a linear dynamics with Langevin noise. This choice was motivated by the fact that the posterior statistics, i.e. the statistics of hidden trajectories conditioned on observations, is Gaussian and thus fully understandable in terms of first and second moments. The posterior means give us the optimal hidden state predictions, and the posterior variances the error bars on these, hence also the expected inference error.

The approximation we make by the Extended Plefka Expansion is to consider the second moments of the posterior as local in the degrees of freedom but maintaining them as functions of two times. Because of this, the method yields effective equations of motion for the hidden variables that treat these as decoupled from each other and from the observed variables but with nontrivial temporal self-correlations through memory integrals and coloured noises. In mathematical terms, the Extended Plefka Expansion gives us a system of coupled equations for the posterior physical correlations, responses and auxiliary correlations, the latter having the role of implementing the constraint arising from knowledge of the dynamics of the observed nodes. Evaluating the posterior covariances at equal times gives the posterior error, hence the Plefka estimate of the inference error.

Having derived the Plefka equations in general form, we then focussed on systems with weak, long-range, random interactions. Because of the mean field character of the decoupling we assumed in setting up the Plefka expansion, one would expect the approximation to become exact for such a system, in the limit of a large network. We can verify this by comparing to the results of a random matrix theory analysis of the exact inference approach (Kalman filter), in the corresponding limit cases, and of a dynamical functional calculation, which will be both the focus of chapter 4. This provides further evidence to indicate the usefulness of the Extended Plefka Expansion as developed in chapter 2, whose application we generalized here to inference problems for a linear stochastic dynamics of continuous variables.

Looking beyond the above exactness results, it is worth stressing that the approach presented here allows one to derive results for more general interaction scenarios. Specifically it holds for any degree of symmetry of the interactions and is thus applicable to non-equilibrium dynamics, for which the detailed balance condition is not satisfied.

In a second part of the chapter we studied the properties of the inference error as a function of the relevant dimensionless parameters $(\alpha, \gamma, p, \eta)$. These parameters are assumed to be known, either by direct measurement or by theoretical estimation, and our predictions for the posterior statistics then quantify their interplay in determining the inference error. As the parameter space is relatively large, we organized the analysis around the critical regions where the (suitably non-dimensionalized) prediction error diverges. We focussed attention on the power spectrum of the posterior correlations, deploying critical scaling approaches to identify the relevant variables and obtain scaling functions that serve as master curves for appropriately scaled numerical data.

The first critical region we analyzed concerns $\alpha \ll 1$, where there are many fewer observed nodes than hidden ones. Here we found that the presence of interaction symmetry ($\eta = 1$) leads to quite different scaling behaviour than for the generic case $-1 < \eta < 1$, indicating the importance of even small deviations from detailed balance for the dynamics. This is in qualitative agreement with earlier studies on systems without observations, e.g. [50]. We were able to study the crossover from equilibrium to generic non-equilibrium dynamics by including $\epsilon = 1 - \eta$ as a small parameter in the scaling analysis.

The second critical region is $0 < \alpha < 1$ and $p \to 0$, where some parts of the hidden dynamics are strongly constrained but because $\alpha < 1$ there are other parts that remain unconstrained as there are still not enough observed nodes. We identified an analogue of the resulting inference error divergence in studies on "underconstrained" learning in neural networks, e.g. [65, 66]. There one finds a divergence when the number of patterns to be learned equals the number of degrees of freedom. This happens when no weight decay is imposed on the dynamics, which in our scenario corresponds to small λ and hence small p.

It will be interesting to understand better the comparison of our findings with equivalent equilibrium problems where these exist. In fact, for symmetric couplings, the stationary regime of the hidden dynamics is effectively at equilibrium. One could thus apply the static Plefka expansion, using means and equal time correlations as order parameters, to the same inference scenario. The main difference would be the fact that the statistics of the hidden values would not be determined from observations over time as in our analysis, but from snapshots of states of the observed nodes. As information on the temporal sequence of observations is lost, this would be expected to lead to a larger inference error. A calculation mirroring ours but in the equilibrium setup could therefore quantify the loss in prediction performance from ignoring the temporal aspect of observational data. This could be an important baseline study establishing for future work the benefits of modelling data explicitly as a trajectory over time.

There are many other directions of future and further investigation. A number of these could start from the data likelihood that the extended Plefka expansion predicts: this is an approximate dynamical action for the observed nodes, having integrated out the hidden node dynamics. As such it captures the observed subnetwork dynamics and a comparison with other approaches tailored to reduced subnetwork descriptions - based on path integrals as well (see chapter 5) or projection methods [77] - should be revealing for further applications in this context: we will sketch it in appendix I, section I.1.1. An additional way of using the Plefka data likelihood would be to learn unknown hidden-to-hidden or observed-to-hidden couplings, which is an important statistical modelling problem for dynamical data. Optimizing the data likelihood with respect to the couplings would by definition give a maximum-likelihood estimate for these quantities. The relevant learning rules could be developed starting from the Plefka equations for fixed couplings. It would be interesting then to compare with related studies for non-equilibrium Ising spins [59] and for networks with binary visible units and continuous-valued hidden ones [69], both also relying on mean field dynamical descriptions. One could investigate interesting questions such as the accuracy of the inferred couplings, e.g. as in [78], and the computational efficiency of the iterative algorithms for implementing the learning rules.

An application of the Extended Plefka Expansion to spin systems might also be worthwhile, both for inference of hidden states, as done in [37], and learning of interactions. The analytically tractable scenario of an infinitely large network of spins with random asymmetric couplings has already been studied by a replica approach [57]; intriguingly, there the error incurred in predicting the states of hidden nodes does not exhibit a singularity structure like to the one presented here.

Finally, exploiting further the analogy with dynamical learning in neural networks, one could think of several additional studies, such as including noise in the observation process [79–82] or investigating finite-size effects [83].

Appendix D

Discrete time operator inversion

D.1 Discrete time operator inversion

One has to exploit operator inversion to work out an expression for the second order statistics of bulk nodes from the path integral representation of the likelihood (3.2.24): the Gaussian structure of its integrand lends itself to simple manipulations in this direction. We neglect for this task terms containing $\psi_i^{\text{eff}}(t)$ and $I_i^{\text{eff}}(t)$ as they act linearly on the $\boldsymbol{x}$s, thus they cause simply a shift of Gaussian means. Times t, t' are first discretized with unit time steps Δ. By construction (see the main text and [23]) $\delta\hat{x}_i(T) = \delta\hat{x}_a(T) = 0$ and we fix $\delta x_i(0) = 0$ to have the same number of integration variables for x_i and $\hat{x}_i$, thus to make all blocks of the covariance and the inverse covariance square.[1] We define the vectors $\delta\boldsymbol{x}_i = \delta x_i(\Delta), ..., \delta x_i(T)$ and $\mathrm{i}\delta\hat{\boldsymbol{x}}_i = \mathrm{i}\delta\hat{x}_i(0), ..., \mathrm{i}\delta\hat{x}_i(T-\Delta)$ - δs indicate deviations from means - and we consider for each i a "generalized" covariance matrix (i.e. it contains correlations with auxiliary variables) of size $2T \times 2T$

$$\mathbf{C}_{i\,\text{gen}} = \left\langle \begin{pmatrix} \delta\boldsymbol{x}_i \\ -\mathrm{i}\delta\hat{\boldsymbol{x}}_i \end{pmatrix} \begin{pmatrix} \delta\boldsymbol{x}_i & -\mathrm{i}\delta\hat{\boldsymbol{x}}_i \end{pmatrix} \right\rangle \tag{D.1.1}$$

[1] A consequence of this assumption is that we would need then to include explicitly a term $\delta(t)\delta(t')\delta C_i(0,0)$ in $\hat{B}_i^{\text{eff}}(t,t')$ to get the desired initial Gaussian distribution (more rigorously, it would be the distribution of $\delta x_i(\Delta)$ and not $\delta x_i(0)$, but such a difference becomes irrelevant for $\Delta \to 0$).

More explicitly it exhibits the following structure

$$\mathbf{C}_{i\,\mathrm{gen}} = \begin{pmatrix} \langle \delta x_i(\Delta)\delta x_i(\Delta)\rangle & \dots & \langle \delta x_i(\Delta)\delta x_i(T)\rangle & -\langle \delta x_i(\Delta)\mathrm{i}\delta\hat{x}_i(0)\rangle & \dots & -\langle \delta x_i(\Delta)\mathrm{i}\delta\hat{x}_i(T-\Delta)\rangle \\ \vdots & & \ddots & \ddots & & \vdots \\ \langle \delta x_i(T)\delta x_i(\Delta)\rangle & \dots & \langle \delta x_i(T)\delta x_i(T)\rangle & \ddots & \dots & -\langle \delta x_i(T)\mathrm{i}\delta\hat{x}_i(T-\Delta)\rangle \\ -\langle \mathrm{i}\delta\hat{x}_i(0)\delta x_i(\Delta)\rangle & \dots & \ddots & \langle \mathrm{i}\delta\hat{x}_i(0)\mathrm{i}\delta\hat{x}_i(0)\rangle & \dots & \langle \mathrm{i}\delta\hat{x}_i(0)\mathrm{i}\delta\hat{x}_i(T-\Delta)\rangle \\ \vdots & & \ddots & \ddots & & \vdots \\ -\langle \mathrm{i}\delta\hat{x}_i(T-\Delta)\delta x_i(\Delta)\rangle & \dots & \dots & \dots & \dots & \langle \mathrm{i}\delta\hat{x}_i(T-\Delta)\mathrm{i}\delta\hat{x}_i(T-\Delta)\rangle \end{pmatrix}$$

The inverse generalized covariance can be read from (3.2.24) and has a block structure of the type

$$\mathbf{C}^{-1}_{i\,\mathrm{gen}} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{pmatrix}$$

$$\mathbf{A}_1 = \begin{pmatrix} -\Delta^2\hat{C}_i^{\mathrm{eff}}(\Delta,\Delta) & -\Delta^2\hat{C}_i^{\mathrm{eff}}(\Delta,2\Delta) & \dots & \dots & -\Delta^2\hat{C}_i^{\mathrm{eff}}(\Delta,T) \\ -\Delta^2\hat{C}_i^{\mathrm{eff}}(2\Delta,\Delta) & -\Delta^2\hat{C}_i^{\mathrm{eff}}(2\Delta,2\Delta) & \dots & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\Delta^2\hat{C}_i^{\mathrm{eff}}(T,\Delta) & \dots & \dots & \dots & -\Delta^2\hat{C}_i^{\mathrm{eff}}(T,T) \end{pmatrix}$$

$$\mathbf{A}_3 = \begin{pmatrix} (1-\Delta^2\hat{R}_i^{\mathrm{eff}}(0,\Delta)) & -\Delta^2\hat{R}_i^{\mathrm{eff}}(0,2\Delta) & \dots & \dots & -\Delta^2\hat{R}_i^{\mathrm{eff}}(0,T) \\ (-1+\Delta\lambda) & (1-\Delta^2\hat{R}_i^{\mathrm{eff}}(\Delta,2\Delta)) & \dots & \dots & \dots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & (-1+\Delta\lambda) & (1-\Delta^2\hat{R}_i^{\mathrm{eff}}(T-\Delta,T)) \end{pmatrix}$$

The bottom left block vanishes because of the causality property of $\hat{R}_i^{\mathrm{eff}}$; clearly $\mathbf{A}_2 = \mathbf{A}_3^T$.

$$\mathbf{A}_4 = \begin{pmatrix} -\Delta^2\left(\hat{B}_i^{\mathrm{eff}}(0,0)+\frac{\Sigma_i}{\Delta}\right) & -\Delta^2\hat{B}_i^{\mathrm{eff}}(0,\Delta) & \dots & \dots & -\Delta^2\hat{B}_i^{\mathrm{eff}}(0,T-\Delta) \\ -\Delta^2\hat{B}_i^{\mathrm{eff}}(\Delta,0) & -\Delta^2\left(\hat{B}_i^{\mathrm{eff}}(\Delta,\Delta)+\frac{\Sigma_i}{\Delta}\right) & \dots & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\Delta^2\hat{B}_i^{\mathrm{eff}}(T-\Delta,0) & \dots & \dots & \dots & -\Delta^2\left(\hat{B}_i^{\mathrm{eff}}(T-\Delta,T-\Delta)+\frac{\Sigma_i}{\Delta}\right) \end{pmatrix}$$

Also $\mathbf{C}_{i\,\mathrm{gen}}$ (D.1.1) has a block structure which can be written

$$\mathbf{C}_{i\,\mathrm{gen}} = \begin{pmatrix} \mathbf{A}'_1 & \mathbf{A}'_2 \\ \mathbf{A}'_3 & \mathbf{A}'_4 \end{pmatrix}$$

To find $\delta C_i(t,t') = \langle \delta x_i(t) \delta x_i(t') \rangle$, $\delta B_i(t,t') = -\langle \delta \hat{x}_i(t) \delta \hat{x}_i(t') \rangle$, $\delta R_i(t',t) = -\langle \mathrm{i} \delta \hat{x}_i(t) \delta x_i(t') \rangle$ (i.e. the elements of the blocks $\boldsymbol{A}'_1$, $\boldsymbol{A}'_2$ and $\boldsymbol{A}'_4$) as functions of the effective fields (contained in $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{A}_4$) one simply uses the fact that $\mathbf{C}_{i\,\mathrm{gen}}$ and $\mathbf{C}^{-1}_{i\,\mathrm{gen}}$ are linked by

$$\sum_{\tau} \mathbf{C}_{i\,\mathrm{gen}}(t,\tau) \mathbf{C}^{-1}_{i\,\mathrm{gen}}(\tau,t') = \mathbb{1} \delta_{tt'} \tag{D.1.2}$$

The matrix product involves summations over the time index τ that, in the limit $\Delta \to 0$, translate into temporal integrals: see equations (3.2.37a), (3.2.37b) and (3.2.37c) (note that there we drop the δs for brevity but all second moments must be intended as connected).

The calculation for the observed nodes is analogous and simpler, as $\mathbf{C}_{a\,\mathrm{gen}}$ reduces to just one type of element $C_{a\,\mathrm{gen}}(t,t') = \delta B_a(t,t')$, as shown in the main text.

Chapter 4

Inference for dynamics: exact average case performance

This work has been done in collaboration with Manfred Opper (TUB Berlin). The related paper has been accepted for publication in Physical Review E [84].

> ὡς εἰκόνι μέν, ἐπείπερ οὐδ' αὐτὸ τοῦτο ἐφ' ᾧ γέγονεν ἑαυτῆς ἐστιν, ἑτέρου δέ τινος ἀεὶ φέρεται φάντασμα, διὰ ταῦτα ἐν ἑτέρῳ προσήκει τινὶ γίγνεσθαι, οὐσίας ἁμωσγέπως ἀντεχομένην, ἢ μηδὲν τὸ παράπαν αὐτὴν εἶναι [1]
>
> *Plato, "Timaeus", 52c*

4.1 Introduction

Inferring the time evolution of a partially observed system of continuous degrees of freedom (d.o.f.) is an important problem in statistical physics, as we have begun to discuss in chapter 3. In systems biology these d.o.f. might for example be concentrations of interacting molecular species in biochemical networks. Inference of unobserved or hidden d.o.f. is then often crucial, e.g. for an understanding of molecular mechanisms underlying genetic and metabolic processes. Hidden d.o.f. can occur because the behaviour of part of a network is simply not recorded, or because the amount of experimental data available might be limited [85]. If as in our analysis one

[1] Namely, that to an image it belongs, seeing that it is not the very model of itself, on which itself has been created, but fleets ever as a phantom of something else, to come into existence in some other thing, clinging to existence as best it may, on pain of being nothing at all.

studies generic continuous d.o.f., a potentially broad and interdisciplinary range of applications can be envisaged beyond biology, e.g. in financial data [86] or weather forecasting [87].

Inference has been studied using statistical mechanics approaches predominantly in scenarios without a temporal dimension, e.g. when learning from examples in neural networks [88, 89]. In all these works, the equilibrium assumption was crucial: learning of weights is intended as a minimization of an energy function, see e.g. [90]. Several studies have, like ours, focussed on performance analysis in the thermodynamic limit of large systems [83, 90]. Especially for linear learning problems, the spectrum of the input correlation matrix (or equivalently the average response function) has turned out to be a key quantity and has been studied by different means, including the replica method [66, 89, 91] based on the pioneering work of Edwards and Jones [92], diagrammatic techniques [65] and partial differential equations from matrix identities [83]. A key system parameter is the "storage" ratio between the number of training examples and the number of parameters to be learned [65, 91].

Rather less work has been done for inference based on entire temporal trajectories, with most efforts focussed on the dynamics of discrete variables, typically Ising spins with random asymmetric couplings: see [37] for a review and [57–59, 93] for examples. We extend these studies significantly by accounting for generic interaction symmetry, thus allowing us to interpolate across a range of non-equilibrium situations all the way to equilibrium dynamics. The results we present here are *exact* in the thermodynamic limit and complement the study of chapter 3, done using an a priori approximate method, the Extended Plefka Expansion. Our emphasis on non-equilibrium dynamics is motivated by the fact that many biological processes are out of equilibrium. Indeed, recent studies [94] and computational models [95] have called for a non-equilibrium approach to gene expression dynamics that would allow one to infer regulatory interactions and transcription factor activity from time-resolved measurements.

As in chapter 3, we focus on a paradigmatic scenario: stochastic linear dynamics on a network of continuous d.o.f. that interact via random Gaussian couplings. Such linear dynamics should give a reasonable account also of the behaviour of generic nonlinear networks of continuous d.o.f. near stable fixed points. We recall that the distribution over network trajectories is Gaussian in this setting, and hence so is the posterior over hidden trajectories given a time trajectory of the observed nodes. Its mean gives the optimal prediction of the time-dependent hidden state, while the second order statistics give information on the certainty of this prediction. In particular, the normalized trace of the equal-time posterior covariance matrix will be our measure of inference error, in continuity with chapter 3. Posterior covariances between different times quantify temporal

correlations of prediction uncertainties.

The choice of a linear dynamics with random couplings is motivated by the fact that, on the one hand, it can be addressed by means of existing algorithmic tools of state space models and, on the other, it can be treated exactly in the limit of infinitely large networks. Indeed in statistical modelling problems as state inference, situations where only a fraction of the system is observable are dealt with by *state space models*, i.e. by the introduction of *hidden* variables [16]. Hidden variables play the role of unobserved states (this is the reason why we refer to them equivalently as "hidden" or "unobserved"), as opposed to observed states, modelling actual data. In particular, if hidden and observed variables are assumed to be drawn from a multivariate Gaussian distribution, we have what is called a linear-Gaussian state space model, which is equivalent to a linear dynamical system describing the coupled temporal evolution of the two types of variables. We show in fact that our setting is closely related to (linear Gaussian) state space modelling in statistics, where the dynamics of hidden variables can only be observed indirectly. This allows us to deploy inference methods developed for such models, specifically the Kalman filter (and smoother) [63]. Generally *filtering* and *smoothing* serve the purpose of reconstructing the state of a system when only some information is available through measurements. These "state prediction" techniques constitute a vast area of enquiry which has grown significantly in the last decades and their presentation goes beyond the scope of this chapter (see e.g. standard textbooks [12, 16, 96]). Kalman Filter equations [63] yield an estimate of the hidden dynamics (conditional on data) and a covariance matrix associated to that particular estimate. The performance is governed by Lyapunov and Riccati recursions, for which theorems of convergence and stability properties are well known [96], in particular for the fixed point equations encoding the steady state solutions.

The novelty of our approach is that we assess the inference error of the Kalman filter for *random* interactions, which induce a random distribution in the eigenvalues of the posterior covariance. In the thermodynamic limit of large networks that we consider, the spectrum becomes self-averaging: its fluctuations tend to zero, and it becomes equal to the disorder (random interaction) average of the spectrum. We tackle this disorder average by exploiting Random Matrix Theory (RMT) results [36]. For related approaches that connect RMT and Bayesian statistics see [97, 98] and also references therein.

We will see that the combination of Kalman filter and RMT gives a wealth of information for inference in systems with equilibrium dynamics, i.e. obeying detailed balance, but cannot be extended in an obvious way to non-equilibrium dynamics. For these scenarios we choose an alternative avenue, using dynamical functionals.

Tools and vocabulary in this second approach are inspired by statistical physics. One can in fact re-phrase an inference problem in a framework akin to thermodynamics, by noting that the likelihood of data resemble a partition function over hidden variables. Finding the average posterior distribution is mapped into a *quenched* average of a free energy w.r.t. to couplings: rigorously one should appeal to a replicas construction, yet for the Gaussian fully connected model under consideration we find that the *annealed* version, while being much simpler to perform, gives equivalent answers. The replica approach was used for inference of spins trajectories in [57] generalizing to dynamics an approach that was already used for learning in static network (see [88–90]).

The aim of this chapter is to provide exact results on the average inference error for large size networks, against which other approximation methods or algorithms can be compared. Exactness in the thermodynamic limit relies crucially on the assumption of weak long-range (mean field) interactions. In addition to the use of Kalman filter recursions combined with RMT, as well as dynamical functionals, we provide a link to variational methods (see in particular appendix F).

The chapter is organized as follows. After presenting the governing Kalman filter equations for the posterior variance and the effective posterior drift (section 4.2 and appendix E), we use RMT to study the equilibrium dynamics case in section 4.3, first for the elementary case of hidden variables with only self-interactions (section 4.3.2), then for symmetric hidden-hidden couplings (section 4.3.3), where we apply free probability methods. Moving on to non-equilibrium dynamics, we describe in section 4.4 the dynamical functional method. We focus on the fully asymmetric case (section 4.4.1) initially, which then generalizes to arbitrary symmetry (section 4.4.2). The result is an algebraic equation for the stationary posterior variance in the Laplace domain which coincides with what we derived by the Extended Plefka Expansion in chapter 3. We summarize and discuss the outlook for future work in section 4.5.

4.2 Model and general expression for posterior covariance

The setting we study consists of two sets of variables: the subnetwork, which models the *observed* d.o.f. and the bulk, which stays *hidden* and whose values we want to infer from the observations. To allow explicit insight into how the inference performance depends on the structural parameters of the problem we consider a tractable scenario, where subnetwork and bulk interact *linearly*.

Our model, then, is a linear dynamical system specified by the following equations

$$\partial_t \boldsymbol{x}^{\mathrm{b}}(t) = \boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{x}^{\mathrm{b}}(t) + \boldsymbol{\xi}^{\mathrm{b}}(t) \tag{4.2.1a}$$

$$\partial_t \boldsymbol{x}^{\mathrm{s}}(t) = \boldsymbol{K}^{\mathrm{ss}}\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{x}^{\mathrm{b}}(t) + \boldsymbol{\xi}^{\mathrm{s}}(t) \tag{4.2.1b}$$

where subnetwork and bulk variables are denoted respectively by the superscript s and b; $\boldsymbol{\xi}^{\mathrm{s}}(t)$ and $\boldsymbol{\xi}^{\mathrm{b}}$ are independent white Gaussian noises with zero mean and variance

$$\langle \boldsymbol{\xi}^{\mathrm{s}}(t)\boldsymbol{\xi}^{\mathrm{s}}(t')^{T}\rangle = \boldsymbol{\Sigma}^{\mathrm{ss}}\delta(t-t') \tag{4.2.2a}$$

$$\langle \boldsymbol{\xi}^{\mathrm{b}}(t)\boldsymbol{\xi}^{\mathrm{b}}(t')^{T}\rangle = \boldsymbol{\Sigma}^{\mathrm{bb}}\delta(t-t') \tag{4.2.2b}$$

In addition the matrix $\boldsymbol{K}^{\mathrm{ss}}$ ($\boldsymbol{K}^{\mathrm{bb}}$) contains the linear couplings between subnetwork (bulk) variables while $\boldsymbol{K}^{\mathrm{bs}}$, $\boldsymbol{K}^{\mathrm{sb}}$ specify the interactions between subnetwork and bulk. This dynamics is completely analogous to the set of starting equations of chapter 3, namely (3.2.1a) and (3.2.1b), provided that one sets $\boldsymbol{K}^{\mathrm{bb}} = \{-\lambda + J_{ij}\}$, $\boldsymbol{K}^{\mathrm{ss}} = \{-\lambda + J_{ab}\}$ and $\boldsymbol{K}^{\mathrm{sb}} = \{K_{ai}\}$ (similarly for $\boldsymbol{K}^{\mathrm{bs}}$).

As pointed out in the introduction, a linear system with Gaussian noise produces a Gaussian distribution over the dynamical trajectories of the entire network. By this we mean that, if one considers a time discretized version of the dynamics (4.2.1a) and (4.2.1b), the joint distribution of the collection of subnetwork and bulk variables across all time steps is Gaussian, as also shown in appendix E. Inferring the hidden dynamics then corresponds to Gaussian conditioning. In particular, the aim is to evaluate the posterior probability distribution over hidden trajectories, conditioned on the observed subnetwork trajectory. We denote the latter $\boldsymbol{X}^{\mathrm{s}}$, as a shorthand for the data sequence $\{\boldsymbol{x}^{\mathrm{s}}(t)|t \in [0,T]\}$. The posterior distribution is then fully characterized by the first and second moments

$$\langle \boldsymbol{x}^{\mathrm{b}}(t)\rangle = \boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t) \tag{4.2.3}$$

$$\langle \delta\boldsymbol{x}^{\mathrm{b}}(t)\delta\boldsymbol{x}^{\mathrm{b}}(t')^{T}\rangle = \boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t') \tag{4.2.4}$$

where $\delta\boldsymbol{x}^{\mathrm{b}}(t) = \boldsymbol{x}^{\mathrm{b}}(t) - \boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t)$ is the deviation from the posterior mean and the T superscript denotes vector or matrix transpose. As defined, $\boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t)$ is then the posterior covariance matrix of $\boldsymbol{x}^{\mathrm{b}}(t)$. We shall drop the superscripts for the sake of brevity so will denote $\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t)$ simply by $\boldsymbol{\mu}(t)$ and $\boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t')$ by $\boldsymbol{C}(t,t')$. The best estimate – in the mean-square sense – of the hidden dynamics based on the observed time series $\boldsymbol{X}^{\mathrm{s}}$ is then just $\boldsymbol{\mu}(t)$, while $\boldsymbol{C}(t,t)$ determines the uncertainty in this prediction: in particular, the trace of $\boldsymbol{C}(t,t)$ is the total mean squared prediction error for the hidden variables. Normalizing by the number of hidden nodes defines what we will call the inference error.

To find the posterior means and variances in linear-Gaussian state models one can use a message passing algorithm known as *Kalman Filter* [63] (see appendix E). The Kalman Filter consists of a *forward* and a *backward* pass. The forward pass iteratively includes observations from time 0 up to t, for increasing t, while the backward pass – also known as "smoothing" – incorporates information from the final time T to t, for decreasing t. For a long time series, the algorithm will

converge to stationary values for the covariances when well away from the two ends $t = 0$ and $t = T$; note though that the state prediction $\boldsymbol{\mu}(t)$ remains time dependent as it is driven by the time dependence of the observed $\boldsymbol{x}^{\mathrm{s}}(t)$. The covariances, on the other hand, are entirely independent of the $\boldsymbol{x}^{\mathrm{s}}(t)$, by a general property of conditional Gaussian distributions: they only depend on which variables are observed, but not their values. Note that this contrasts with the case of e.g. binary spins, where mean and variance are directly related so that also variances of individual spins would generally be non-stationary.

The stationary inference error, i.e. the normalized trace of the stationary equal time posterior covariance $\boldsymbol{C}(t,t) = \boldsymbol{C}$, will be the main focus of our attention. As shown in appendix E, $\boldsymbol{C}$ satisfies

$$\boldsymbol{K}^{\mathrm{bb|s}}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\mathrm{bb|s}\,T} + \boldsymbol{\Sigma}^{\mathrm{bb}} = 0 \tag{4.2.5}$$

This is a Lyapunov equation with an “effective" or “posterior" drift $\boldsymbol{K}^{\mathrm{bb|s}}$, where we use the superscript bb|s to indicate that this is the bulk-bulk coupling matrix conditioned on the observed subnetwork trajectory. By “posterior" we mean then that $\boldsymbol{K}^{\mathrm{bb|s}}$ incorporates the effect of the observations and defines an effective posterior dynamics

$$\partial_t \delta\boldsymbol{x}^{\mathrm{b}}(t) = \boldsymbol{K}^{\mathrm{bb|s}}\delta\boldsymbol{x}^{\mathrm{b}}(t) + \boldsymbol{\xi}^{\mathrm{b}}(t) \tag{4.2.6}$$

The effective drift can be written as

$$\boldsymbol{K}^{\mathrm{bb|s}} = \boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A} \tag{4.2.7}$$

where $\boldsymbol{A} = \boldsymbol{A}^T$ is a symmetric matrix that is a solution of the matrix Riccati (i.e. quadratic) equation

$$\boldsymbol{A}\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb}\,T}\boldsymbol{A} = \boldsymbol{W} \tag{4.2.8}$$

Here the *feedback* matrix $\boldsymbol{W} = \boldsymbol{K}^{\mathrm{sb}\,T}(\boldsymbol{\Sigma}^{\mathrm{ss}})^{-1}\boldsymbol{K}^{\mathrm{sb}}$ describes how observations affect the inferred statistics. This matrix is determined by the interplay between the strength of hidden-observed interactions $\boldsymbol{K}^{\mathrm{sb}}$ and the dynamical noise on the observed variables, namely $\boldsymbol{\Sigma}^{\mathrm{ss}}$. (We stress here that this is noise acting on the time evolution of $\boldsymbol{x}^{\mathrm{s}}$, *not* noise affecting our measurement of the observed trajectory.)

The matrix $\boldsymbol{A}$ in (4.2.7) is directly related to the backwards messages sent in the Kalman filter method. Specifically, the distribution of $\delta\boldsymbol{x}^{\mathrm{b}}(t)$ conditioned only on observations *from time t onwards* is Gaussian, and $\boldsymbol{A}$ is its inverse covariance in the stationary regime. Accordingly, equation (4.2.8) can be derived as the stationary limit of what is known as a Riccati recursion, for the backward pass in the Kalman Filter (see appendix E). Without observations the distribution of

$\boldsymbol{x}^{\mathrm{b}}(t)$ conditional only on data beyond t is flat, hence $\boldsymbol{A}$ vanishes. Then $\boldsymbol{K}^{\mathrm{bb|s}}$ reduces to $\boldsymbol{K}^{\mathrm{bb}}$ as expected and the posterior covariance to the unconditional covariance because (4.2.5) becomes simply $\boldsymbol{K}^{\mathrm{bb}}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\mathrm{bb}\,T} + \boldsymbol{\Sigma}^{\mathrm{bb}} = 0$. One sees therefore that $\boldsymbol{A}$ is the key quantity that captures the effects of the observations on the (second order) posterior statistics. This insight is supported by an alternative variational derivation of (4.2.5), (4.2.7) and (4.2.8), outlined in Appendix F, where $\boldsymbol{A}$ appears as a Lagrange multiplier implementing the constraints resulting from the observed data.

Once the stationary equal-time covariance $\boldsymbol{C}$ has been found, it is clear from (4.2.6) that the two-time covariance must be given by

$$\boldsymbol{C}(t - t') = e^{\boldsymbol{K}^{\mathrm{bb|s}}(t-t')}\boldsymbol{C} \tag{4.2.9}$$

for $t > t'$. This exponential decay with the effective drift matrix $\boldsymbol{K}^{\mathrm{bb|s}}$ can be derived explicitly by generalizing the filtering-smoothing procedure (see appendix E and references there). We have emphasized in the notation the fact that $\boldsymbol{C}(t - t')$ depends only on the time difference because the stationary regime obeys time-translation invariance. Stability of the conditional hidden dynamics, where (4.2.9) decays to zero as $t - t'$ grows, requires $\boldsymbol{K}^{\mathrm{bb|s}}$ to be negative definite. Assuming that the dynamical matrix $\boldsymbol{K}^{\mathrm{bb}}$ of the isolated hidden dynamics has this property, then also $\boldsymbol{K}^{\mathrm{bb|s}}$ does because $\boldsymbol{A}$, as the inverse covariance matrix in the stationary backwards messages, is non-negative definite. As a result, the conditional covariance can be written effectively as a marginal one whose amplitude and decay rate are renormalized by observations. An analogous way of representing posteriors has already turned out to be convenient to generalize existing approximations, such as the Van Kampen system size expansion [99], to the case with observations.

So far in this section we have derived expressions for $\boldsymbol{C}$ and $\boldsymbol{C}(t - t')$ that specify the second order posterior statistics in our setting of inferring hidden state trajectories. These results are valid for given values of the interaction matrices $\boldsymbol{K}^{\mathrm{bb}}$ etc. In the remainder of the chapter we consider these interactions to be drawn from some probability distribution, acting as *quenched disorder*. In an appropriately defined infinite size or thermodynamic limit we then expect key results such as the eigenvalue spectrum of $\boldsymbol{C}$ to be self-averaging, i.e. independent of the specific realization. In particular we look at a fully connected system interacting via Gaussian couplings. This is a standard scenario used to analyze the mean-field regime of e.g. spin glass models [50]. It can also be thought as the large connectivity limit of an Erdős-Rényi graph [100] with Gaussian weights; studying dynamical processes on such random graphs to predict the evolution of each node from partial observations is of interest in e.g. epidemic forecasting [101, 102].

A precedent for the use of RMT techniques, such as Stieltjes transforms and free probability, in the study of asymptotic eigenvalue distributions for random Lyapunov and Riccati recursions –

like those occurring in filtering – can be found in [97]. Ref. [97] takes a control and systems theory perspective, however, while we focus on inference for dynamics. It is worth stressing that this makes our approach more general, as we look at a time dependent problem with a quenched, "frozen" randomness rather than a sequence of signals where the randomness could be re-sampled at each step. From the spectrum $\boldsymbol{C}$ we will obtain the inference error; we will also study the properties of the posterior drift $\boldsymbol{K}^{\mathrm{bb|s}}$, whose inverse defines the spectrum of relaxation times of the posterior dynamics.

4.3 Thermodynamic limit by Random Matrix Theory

To investigate the thermodynamic limit, we first apply tools from random matrix theory (RMT) to *equilibrium* dynamics, where detailed balance holds. We study two such scenarios. In the first, the hidden variables only have self-interactions (Sec. 4.3.2); in the second we add random symmmetric hidden-to-hidden interactions (Sec. 4.3.3). In both cases we make the same assumptions regarding the hidden-to-observed interactions $\boldsymbol{K}^{\mathrm{sb}}$, and therefore discuss first the resulting statistics of the feedback matrix $\boldsymbol{W}$.

4.3.1 Feedback matrix: Wishart ensemble

The feedback matrix $\boldsymbol{W} = \boldsymbol{K}^{\mathrm{sb}\,T}(\boldsymbol{\Sigma}^{\mathrm{ss}})^{-1}\boldsymbol{K}^{\mathrm{sb}}$ is a positive definite symmetric matrix of size $N^{\mathrm{b}} \times N^{\mathrm{b}}$, where N^{b} is the number of hidden variables, i.e. the number of components of the vector $\boldsymbol{x}^{\mathrm{b}}$. We assume throughout in the following that the elements of the $N^{\mathrm{s}} \times N^{\mathrm{b}}$ matrix $\boldsymbol{K}^{\mathrm{sb}}$ are independent zero mean Gaussian random variables of fixed variance. If $\boldsymbol{\Sigma}^{\mathrm{ss}} = \sigma_{\mathrm{s}}^2\mathbb{1}$ is isotropic, $\boldsymbol{W}$ is then a sample from a *Wishart* random matrix ensemble, whose spectral properties are well understood [36]. In the thermodynamic limit of infinitely large matrices, $N^{\mathrm{b}} \to \infty$, and up to an overall scale of the eigenvalues, the eigenvalue density of $\boldsymbol{W}$ is thus given by the *Marčenko-Pastur* law (MP) [103]

$$\rho_\alpha(\hat{w}) = (1-\alpha)\Theta(1-\alpha)\delta(\hat{w}) + f_\alpha(\hat{w}) \tag{4.3.1}$$

where

$$f_\alpha(\hat{w}) = \frac{1}{2\pi\hat{w}}\sqrt{(\hat{w}-\hat{w}_-)(\hat{w}_+-\hat{w})} \tag{4.3.2}$$

and is to be read as nonzero only when $\hat{w}$ lies in the interval $[\hat{w}_-, \hat{w}_+]$ with $\hat{w}_\pm = (\sqrt{\alpha} \pm 1)^2$. The delta peak at $\hat{w} = 0$ in (4.3.1) contributes only when $\alpha < 1$, as indicated by the Heaviside step function $\Theta(\cdot)$. Similarly to chapter 3, here we have defined $\alpha = N^{\mathrm{s}}/N^{\mathrm{b}} = N^{\mathrm{observed}}/N^{\mathrm{hidden}}$ as the fundamental parameter of our analysis, giving the ratio and thus the relative importance of the

sizes of the observed and unknown "sectors" of our network. Let us also recall that this parameter resembles the storage ratio [88, 90], or number of training examples per parameter to be learned, in neural network learning. Indeed, in the context of learning linear relationships from examples, the distribution (4.3.1) also gives the spectrum of the input correlation matrix governing the learning dynamics [65, 66, 83, 91].

In the spectrum (4.3.1) the δ peak at $\hat{w} = 0$ arises from the $N^{\mathrm{b}} - N^{\mathrm{s}} = N^{\mathrm{b}}(1-\alpha)$ directions in the hidden state space that are not directly constrained by observations when $\alpha < 1$. The remaining $f_\alpha(\hat{w})$ piece is a semi-circle in the interval $[\hat{w}_-, \hat{w}_+]$, distorted by a factor $1/\hat{w}$. For $\alpha > 1$ this is the only contribution; in the limit $\alpha \gg 1$ the relative variance of the eigenvalues around their mean $\langle \hat{w} \rangle = \alpha$ goes to zero.

4.3.2 Self-interacting hidden variables

Inference error and relaxation times

We assume below that the noise acting on bulk variables is isotropic, $\boldsymbol{\Sigma}^{\mathrm{bb}} = \sigma_{\mathrm{b}}^2 \mathbb{1}$, as already assumed for the subnetwork noise. This is equivalent to assuming that the amplitude of fluctuations is homogeneous within the hidden system, as it would be if it was given by a physical temperature. Anisotropies would add non-trivial correlations between d.o.f. that would obscure the effect of interactions, which is our main interest here. In this section we restrict the focus further to interactions between bulk and subnetwork, by taking $\boldsymbol{K}^{\mathrm{bb}} = -\lambda \mathbb{1}$ where the self-interaction λ is the only interaction among hidden variables. Given this, any interesting behaviour has to come from observations.

By simultaneously diagonalizing $\boldsymbol{W}$ and $\boldsymbol{A}$, (4.2.8) reduces to a scalar equation relating the eigenvalues of these matrices, respectively w and a, as

$$\sigma_{\mathrm{b}}^2 a^2 + 2\lambda\, a = \frac{k^2}{\sigma_{\mathrm{s}}^2} \hat{w} \tag{4.3.3}$$

where we have extracted from w an amplitude factor by writing $w = k^2 \hat{w}/\sigma_{\mathrm{s}}^2$, k being the amplitude for $\boldsymbol{K}^{\mathrm{sb}}$ entries and $\hat{w}$ a dimensionless Wishart random variable. The physical solution[2] for a is

$$a = \frac{-\lambda + \sqrt{\lambda^2 + \sigma^2 \hat{w}}}{\sigma_{\mathrm{b}}^2} \tag{4.3.4}$$

[2] In principle we get 2 solutions with opposite sign, the choice between them is made by requiring that the limit $\hat{w} \to 0$ is equal to 0, so that in the absence of observations a vanishes (it plays the role of a Lagrange multiplier) and the posterior drift and covariance become ordinary ones.

with the shorthand $\sigma = \sigma_b k/\sigma_s$. By diagonalizing (4.2.7) one then gets for the eigenvalues of $\boldsymbol{K}^{bb|s}$, which we denote by r

$$r = -\lambda - a\,\sigma_b^2 = -\sqrt{\lambda^2 + \sigma^2 \hat{w}} \tag{4.3.5}$$

From (4.2.6) and (4.2.9), the distribution of $-r$ gives the relaxation rate spectrum of the posterior dynamics, and (4.3.5) shows that these rates are increased by observations, i.e. correlations get shorter in time. As expected this effect becomes stronger as the hidden-observed interaction amplitude k increases, at fixed ratio σ_b/σ_s. The stability of the conditional dynamics (i.e. $r < 0$) is trivially satisfied unless $\lambda/\sigma \to 0$ and $\alpha \leq 1$ (the values of α for which 0 eigenvalues of $\hat{w}$ exist): this corresponds to one of the divergence regions we discussed in chapter 3, from section 3.3.

From (4.3.5) we can now find the spectrum of r as the appropriate transformation of the MP law

$$\rho(r) = (1-\alpha)\Theta(1-\alpha)\delta(r+\lambda) + f(\hat{w}(r))|\hat{w}'(r)| \tag{4.3.6}$$

where $f(\hat{w}(r))$ is defined only between $r_\pm = \sqrt{\sigma^2(\sqrt{\alpha} \pm 1)^2 + \lambda^2}$ and $\hat{w}(r) = -(r^2+\lambda^2)/\sigma^2$ is the inverse function of (4.3.5). The first piece, a δ-function at $r = -\lambda$, describes the behaviour for hidden state space directions unconstrained by observations (hidden self-interactions shift the MP spectrum).

The above result for the spectrum can also be expressed as a spectrum $\rho(\tau) = \rho(r)/\tau^2$ of relaxation times $\tau = -1/r$ for the posterior dynamics, for which

$$\rho(\tau) = (1-\alpha)\Theta(1-\alpha)\delta\left(\tau - \frac{1}{\lambda}\right) + f(\hat{w}(\tau))|\hat{w}'(\tau)| \tag{4.3.7}$$

This relaxation time connected to the conditional dynamics gives the temporal correlations between inferred hidden values. We sometimes plot $\rho(\ln\tau) = \tau\rho(\tau)$ to show the full range of τ; this $\ln\tau$-spectrum is the same as the one of $\ln r$ up to a sign change, with spectral edges at $\tau_\pm = -1/r_\mp$ - see figure 4.1 (left).

The long-time ($t - t' \gg 1$) behaviour of the posterior covariance is an exponential decay whose characteristic time can be defined in different ways. The slowest relaxation time is $\tau_{max} = 1/r_{min}$, where r_{min} is the minimum eigenvalue of $\boldsymbol{K}^{bb|s}$

$$r_{min} = \sqrt{\lambda^2 + \sigma^2 \hat{w}_{min}} = \begin{cases} \lambda & \alpha \leq 1 \\ \sqrt{\lambda^2 + \sigma^2(\sqrt{\alpha}-1)^2} & \alpha > 1 \end{cases} \tag{4.3.8}$$

We see the spectrum edge r_- captures the fastest decay rate for the posterior variance when $\alpha > 1$ and for $\alpha \gg 1$ one has $\tau_{max} \sim 1/(\sigma\sqrt{\alpha})$. One can also look at a relaxation time defined as the average over the spectrum $\rho(\tau)$, i.e. $\langle\tau\rangle = \int d\tau\rho(\tau)\,\tau$. Or finally one can consider a root mean

square correlation decay time

$$\tau^{*\,2} = \frac{\int_{-\infty}^{+\infty} t^2 C(t)dt}{2\tilde{C}(0)} = -\frac{1}{2\tilde{C}(0)} \frac{d^2\tilde{C}(\mathrm{i}\omega)}{d^2\omega}\bigg|_{\omega=0} \tag{4.3.9}$$

where the power spectrum $\tilde{C}(\mathrm{i}\omega)$ is obtained by setting $z = \mathrm{i}\omega$ in the Laplace transform (see equation (4.3.18) below) of the correlator $C(t-t') = \mathrm{Tr}\,\boldsymbol{C}(t-t')$ (trace normalized by N^{b}). It is easy to verify that all three relaxation times exhibit the same asymptotic decay $\sim 1/(\sigma\sqrt{\alpha})$ for large α. In figure 4.2 (left) we show a comparison at smaller α. With only few observations, all measures of posterior correlation time are close to the $\alpha = 0$ value $1/\lambda$ while for $\alpha > 1$ they start decreasing, crossing over to the $1/\sqrt{\alpha}$ large α tail; $\tau_{\max}$ shows the least smooth transition between these two regimes. We can summarize the behaviour by saying that with more observations the posterior fluctuations (or error bars on the inferred means) become less correlated in time as predictions become more "tied" to the data observed at any specific moment. This effect is seen in more detail in figure 4.1 (right) where with increasing α the relaxation time spectrum becomes more peaked and shifts towards shorter times.

The posterior covariance matrix $\boldsymbol{C}$ has the same set of eigenmodes as $\boldsymbol{K}^{\mathrm{bb}} = -\lambda\mathbb{1}$ in the current scenario because in (4.2.5) all matrices can be simultaneously diagonalized. The eigenvalues C of $\boldsymbol{C}$ give the posterior variance for each mode, which from (4.2.5) is related to r or τ by

$$C = -\frac{\sigma_{\mathrm{b}}^2}{2r} = \frac{\sigma_{\mathrm{b}}^2}{2}\tau = \frac{\sigma_{\mathrm{b}}^2}{2\sqrt{\lambda^2 + \sigma^2\hat{w}}} \tag{4.3.10}$$

This shows that C decreases with increasing feedback values $\hat{w}$: observations increase prediction accuracy as they should. Because $C \propto \tau$, the above results for the spectrum of τ also apply to that of C; see figures 4.1 and 4.2 (left). For large α in particular the spectrum of C becomes a narrow peak around the asymptotic inference error $C \approx \sigma_{\mathrm{b}}^2/(\sigma\sqrt{\alpha})$: as the amount of observed data grows, the uncertainty on the prediction C shrinks with a $1/\sqrt{\alpha}$ law.

We note as an aside that from the proportionality $C \propto \tau$ one can show that the relaxation time τ^* defined in (4.3.9) can be written in terms of spectral averages as

$$\tau^* = \sqrt{\frac{\langle\tau^4\rangle}{\langle\tau^2\rangle}}. \tag{4.3.11}$$

We write explicitly the integrals in (4.3.9)

$$\int_{-\infty}^{+\infty} t^2 C(t)dt = -\frac{\sigma_{\mathrm{b}}^2}{2}\int_{-\infty}^{+\infty} t^2 \int d\hat{w}\rho(\hat{w})\frac{e^{rt}}{r} = 2\sigma_{\mathrm{b}}^2 \int d\tau\rho(\tau)\tau^4 \tag{4.3.12}$$

where we have used the expression for $C(t)$ we are deriving in the next section (see (4.3.16)), to be averaged w.r.t. $\hat{w}$ MP distribution $\rho(\hat{w})$, and $\tau = -1/r = 1/\sqrt{\lambda^2 + \sigma^2\hat{w}}$. Similarly

$$\tilde{C}(0) = \int_{-\infty}^{+\infty} C(t)dt = \sigma_{\mathrm{b}}^2 \int d\tau\rho(\tau)\tau^2 \tag{4.3.13}$$

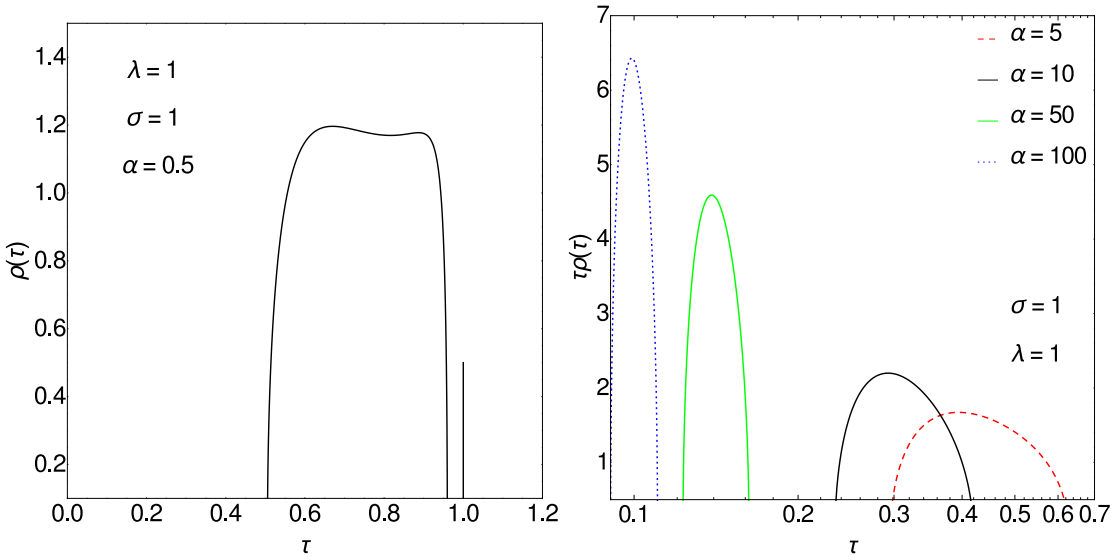

Figure 4.1: (Left) Spectral density $\rho(\tau)$ for $\alpha = 0.5$: the vertical line indicates the δ-peak of height $1 - \alpha$ at $\tau = 1/\lambda$, the relaxation time in the absence of observations. (Right) Spectral density $\rho(\ln \tau) = \tau\rho(\tau)$ of $\ln \tau$: this shifts to smaller $\ln \tau$ as α increases, indicating shorter posterior correlation times. The spectrum also narrows and becomes concentrated around $\tau = 1/\sigma\sqrt{\alpha}$ for large α. As the posterior variance $C \propto \tau$ for each hidden space mode, the distributions of $\ln C$ only differ from those of $\ln \tau$ by a horizontal shift. We shall recall that both τ and $C \sim -1/r$, thus the spectrum of the inference error $\rho(C)$ behaves similarly to $\rho(\tau)$.

By developing the integrals in (4.3.12) and using (4.3.13), one has (4.3.11). Because $\langle\tau\rangle^2\langle\tau^2\rangle \leq \langle\tau^4\rangle$, this implies generally $\langle\tau\rangle \leq \tau^*$ in agreement with the results in figure 4.2 (left).

We finally notice that the proper limits for $\hat{w} \to 0$ (i.e. the results in absence of observations) can be easily retrieved

$$\lim_{\hat{w}\to 0} a = 0 \qquad\qquad \lim_{\hat{w}\to 0} r = -\lambda \tag{4.3.14}$$

from which

$$\lim_{\hat{w}\to 0} \tau = \frac{1}{\lambda} \qquad\qquad \lim_{\hat{w}\to 0} C = \frac{\sigma_{\mathrm{b}}^2}{2\lambda} \tag{4.3.15}$$

as expected.

Posterior covariance in Laplace space

We next turn to the temporal dependence of the posterior covariance (4.2.9). Its trace, normalized by N^{b}, is an average of the contributions from the different eigenmodes of $\boldsymbol{K}^{\mathrm{bb|s}}$. In terms of the relevant eigenvalues $\hat{w}$ and using (4.3.10) these are

$$C_{\hat{w}}(t - t') = e^{r|t-t'|}C = -\frac{\sigma_{\mathrm{b}}^2}{2r}e^{r|t-t'|} \tag{4.3.16}$$

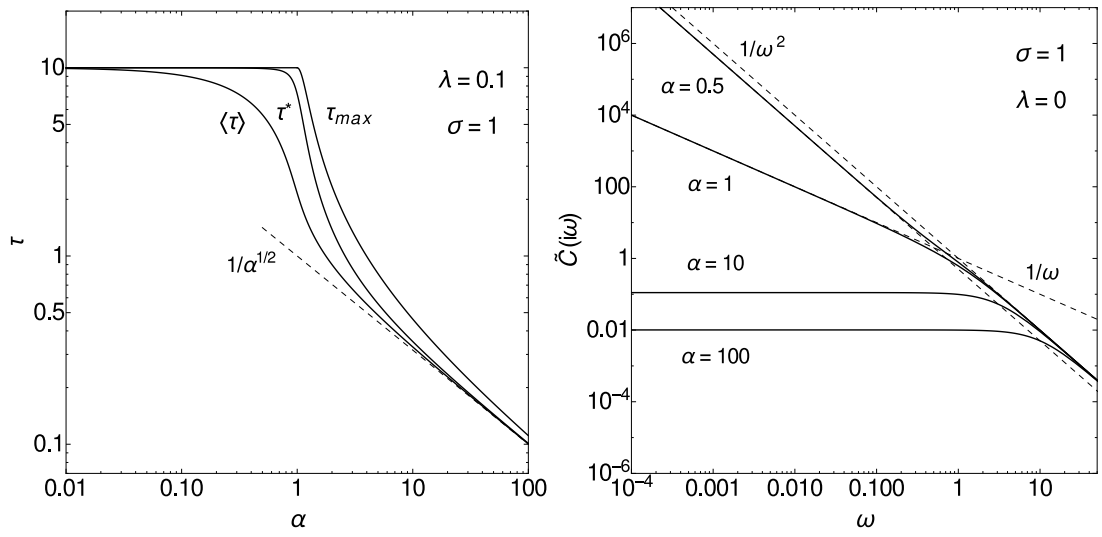

Figure 4.2: (Left) Characteristic posterior relaxation time τ as a function of α, for $\lambda = 0.1$ and $\sigma = 1$, defined in three different ways (see text). For $\alpha \to 0$ all three curves approach $\tau = 1/\lambda = 10$; asymptotically they decay as $1/\sqrt{\alpha}$. In all these regimes, $\tau_{\max}$ describes the slowest relaxation. (Right) Posterior power spectrum (obtained by setting $z = \mathrm{i}\omega$ in (4.3.18)) for various α, at $\lambda = 0$. The power spectrum diverges as $\omega \to 0$ when $\alpha \leq 1$. For small α the divergence is $\propto 1/\omega^2$, crossing over to $\propto 1/\omega$ as $\alpha \to 1$. Beyond $\omega \sim O(1)$ the curves for all α exhibit a standard Lorentzian tail $1/\omega^2$. See chapter 3 for a derivation of these power laws.

with an added subscript $\hat{w}$ to indicate this is the contribution from a single eigenmode, characterized by a specific value of $\hat{w}$. We take the double-sided Laplace transform

$$\tilde{C}_{\hat{w}}(z) = \frac{\sigma_{\mathrm{b}}^2}{2r}\int_{-\infty}^{+\infty} e^{-(z+r)|t'-t|}dt' = \frac{\sigma_{\mathrm{b}}^2}{r^2-z^2} = \frac{\sigma_{\mathrm{b}}^2}{\lambda^2 - z^2 + \sigma^2\frac{k^2}{\sigma_{\mathrm{s}}^2}\hat{w}} = \frac{\sigma_{\mathrm{s}}^2}{k^2}\frac{1}{\frac{\lambda^2-z^2}{\sigma^2}+\hat{w}} \tag{4.3.17}$$

where we have substituted (4.3.5) for r in terms of the self-interaction λ and the feedback matrix eigenvalues $k^2\hat{w}/\sigma_{\mathrm{s}}^2$.

In the thermodynamic limit, we can then get the Laplace transform of the overall covariance normalized trace $C(t-t') = \mathrm{Tr}\,\boldsymbol{C}(t-t')$ by averaging over the Marčenko-Pastur spectrum $\rho(\hat{w})$, yielding $\tilde{C}(z) = \int \tilde{C}_{\hat{w}}(z)\rho(\hat{w})d\hat{w} =$

$$= \begin{cases} \frac{\sigma_{\mathrm{s}}^2}{k^2}(1-\alpha)\frac{\sigma^2}{\lambda^2-z^2} + \frac{1}{2}\frac{\sigma^2}{(\lambda^2-z^2)}\left\{\alpha - 1 - (\frac{\lambda^2-z^2}{\sigma^2}) + \sqrt{[1-\alpha-(\frac{\lambda^2-z^2}{\sigma^2})]^2 + 4(\frac{\lambda^2-z^2}{\sigma^2})}\right\} & \alpha < 1 \\ \frac{\sigma_{\mathrm{s}}^2}{2k^2}\frac{\sigma^2}{(\lambda^2-z^2)}\left\{1 - \alpha - (\frac{\lambda^2-z^2}{\sigma^2}) + \sqrt{[1-\alpha-(\frac{\lambda^2-z^2}{\sigma^2})]^2 + 4(\frac{\lambda^2-z^2}{\sigma^2})}\right\} & \alpha \geq 1 \end{cases}$$

which can be summarized, for each α, as

$$\tilde{C}(z) = \frac{\sigma_{\mathrm{s}}^2}{2k^2}\frac{\sigma^2}{(\lambda^2-z^2)}\left\{1-\alpha-\left(\frac{\lambda^2-z^2}{\sigma^2}\right) + \sqrt{\left[1-\alpha-\left(\frac{\lambda^2-z^2}{\sigma^2}\right)\right]^2 + 4\left(\frac{\lambda^2-z^2}{\sigma^2}\right)}\right\} \tag{4.3.18}$$

The whole expression reduces to a pure Lorentzian $\sigma_b^2/(\lambda^2 - z^2)$ for $\alpha = 0$, i.e. without observations, as expected for a pure exponential decay driven by λ.

One can verify that $\tilde{C}(0)$ has a divergence for $\lambda/\sigma \to 0$ and $\alpha \leq 1$ and the small α-curves in figure 4.2 (right) illustrate this effect: in chapter 3 we provided a systematic study of the approach to such divergences.

4.3.3 Symmetric hidden-hidden couplings

In this section we generalize the above scenario by assuming that $\boldsymbol{K}^{\mathrm{bb}} = -\lambda\mathbb{1} + \boldsymbol{J}$. Here the matrix $\boldsymbol{J}$ provides explicit hidden-to-hidden interactions beyond the self-interaction term $-\lambda\mathbb{1}$ we have had so far. To ensure stability of the hidden system, one requires $\lambda > \lambda_c$ where λ_c is the largest eigenvalue of $\boldsymbol{J}$.

We assume that $\boldsymbol{J}$ is symmetric, which is required for any steady state of the whole system to be at equilibrium, i.e. to obey detailed balance. (One would need to choose the remaining matrices to actually have equilibrium: $\boldsymbol{K}^{\mathrm{ss}}$ symmetric as well as $\boldsymbol{K}^{\mathrm{bs}T} = \boldsymbol{K}^{\mathrm{sb}}$). The posterior drift $\boldsymbol{K}^{\mathrm{bb|s}}$ from (4.2.7) is then also a symmetric matrix. This is crucial as it allows one to solve (4.2.5) and (4.2.8) in closed form. Eq. (4.2.5) gives

$$\boldsymbol{C} = -\frac{\sigma_b^2}{2}\Big((-\lambda + \boldsymbol{J}) - \sigma_b^2\boldsymbol{A}\Big)^{-1} = -\frac{\sigma_b^2}{2}(\boldsymbol{K}^{\mathrm{bb|s}})^{-1} \tag{4.3.19}$$

which is positive definite because $\boldsymbol{K}^{\mathrm{bb|s}} = (-\lambda + \boldsymbol{J}) - \sigma_b^2\boldsymbol{A}$ is negative definite. To eliminate the unknown $\boldsymbol{A}$, note from (4.2.8) that

$$\begin{aligned}\Big((-\lambda + \boldsymbol{J}) - \sigma_b^2\boldsymbol{A}\Big)^2 &= (-\lambda + \boldsymbol{J})^2 + \sigma_b^4\boldsymbol{A}^2 - \sigma_b^2(-\lambda + \boldsymbol{J})\boldsymbol{A} - \boldsymbol{A}(-\lambda + \boldsymbol{J})\sigma_b^2 = \\ &= (-\lambda + \boldsymbol{J})^2 + \sigma_b^2\boldsymbol{W} \doteq \boldsymbol{M}\end{aligned} \tag{4.3.20}$$

where the last equality defines $\boldsymbol{M}$. Hence

$$\boldsymbol{C} = \frac{\sigma_b^2}{2}\Big((-\lambda + \boldsymbol{J})^2 + \sigma_b^2\boldsymbol{W}\Big)^{-\frac{1}{2}} = \frac{\sigma_b^2}{2}\boldsymbol{M}^{-1/2} \qquad \boldsymbol{K}^{\mathrm{bb|s}} = -\boldsymbol{M}^{1/2} \tag{4.3.21}$$

where $\boldsymbol{M}^{1/2}$ is the positive definite square root of $\boldsymbol{M}$ and $\boldsymbol{M}^{-1/2}$ its inverse.

Free probability

From (4.3.21), the spectrum of $\boldsymbol{M}$ directly determines those of $\boldsymbol{C}$ and $\boldsymbol{K}^{\mathrm{bb|s}}$. As a paradigmatic example where this spectrum can be obtained in the thermodynamic limit we consider the case where the elements of $\boldsymbol{J}$ are independently drawn from a Gaussian distribution, i.e. we set $\boldsymbol{J} = j\hat{\boldsymbol{J}}$ with $\hat{\boldsymbol{J}}$ a random matrix from the *Wigner* ensemble [36]. From the Wigner semi-circular law this

has largest eigenvalue 2, thus $\lambda_c = 2j$. We will write the feedback matrix as in section 4.3.2: $\boldsymbol{W} = \frac{k^2}{\sigma_s^2}\hat{\boldsymbol{W}}$ with $\hat{\boldsymbol{W}}$ from the *Wishart* ensemble.

With the above assumptions, $\boldsymbol{M} = (-\lambda + \boldsymbol{J})^2 + \sigma_b^2 \boldsymbol{W}$ is a sum of two independently drawn, symmetric random matrices with known spectrum. Its spectrum can then be found using *free probability* theory. Reviews can be found in [104] for the theory and [105, 106] for applications to RMT. Given two Hermitian random matrices in the infinite size limit, we call them *free* if they are independent and invariant under the random rotation (technically called Haar transformation) that defines free addition [104]: with free matrices, the free and the ordinary addition are equivalent. The sum defining $\boldsymbol{M}$ is effectively a *free* addition in the sense that, because of independent sampling, the eigenvector bases of the two matrices in the sum are randomly rotated against each other. It then turns out that the spectrum of the sum depends only on the eigenvalues and not the eigenvectors of the individual matrices, thus the spectral density of $\boldsymbol{M}$ can be determined exclusively from the one of $\boldsymbol{J}$ and $\boldsymbol{W}$. The intuition beyond this is that, in the limit of infinite matrix size, the detailed statistics of eigenvalues, e.g. whether they are correlated or not, can be neglected [106]. These properties allow one to draw several analogies between free probability for invariant random matrices and classical probability for independent random variables. In particular, while in an ordinary sum of independent random variables it is the cumulants that add, in a free sum of two random matrices it is the R-transforms that are additive [104], and this allows the spectrum of the sum to be determined.

The R transform of a random matrix is related to its Green's function by

$$G(z) = \frac{1}{z - R(G(z))} \tag{4.3.22}$$

The Green's function or resolvent, in turn, is defined for a generic random matrix $\boldsymbol{M}$ as the normalized trace $G_M(z) = \mathrm{Tr}(z - \boldsymbol{M})^{-1}$. It can be written in terms of the eigenvalue density $\rho(m)$ as

$$G_M(z) = \int \frac{\rho(m)}{z - m} dm = \mathrm{Tr}[z - \boldsymbol{M}]^{-1} \tag{4.3.23}$$

which is also known as a Stieltjes transform. Conversely, $\rho(m)$ can be retrieved from the Green's function via

$$\rho(m) = -\frac{1}{\pi} \lim_{\epsilon \to 0^+} \mathrm{Im}\, G_M(m + \mathrm{i}\epsilon) \tag{4.3.24}$$

The route to finding the spectrum of $\boldsymbol{M}$ in our case is then clear: we need to write down the Green's functions and associated R-transforms of $(-\lambda + \boldsymbol{J})^2$ and $\sigma_b^2 \boldsymbol{W}$, respectively, add these two R-transforms to obtain the R-transform of M, and then work backwards to $G_M(z)$ and finally $\rho(m)$.

We denote by $G_1(z)$ the Green's function of $(-\lambda + \boldsymbol{J})^2$, which is given by the integral

$$G_1(z) = \int \frac{\rho(\hat{\jmath})}{z-(-\lambda+j\hat{\jmath})^2} d\hat{\jmath} = \int_{-2}^{2} \frac{\sqrt{4-\hat{\jmath}^2}}{2\pi} \frac{1}{z-(-\lambda+j\hat{\jmath})^2} d\hat{\jmath} \tag{4.3.25}$$

where the Wigner semicircular law has been used. The integral can be performed in closed form

$$G_1(z) = \frac{1}{2j^2} - \frac{1}{4j^2}\sqrt{\frac{(\lambda-\sqrt{z})^2-4j^2}{z}} - \frac{1}{4j^2}\sqrt{\frac{(\lambda+\sqrt{z})^2-4j^2}{z}} \tag{4.3.26}$$

and (4.3.22) then gives the R-transform

$$R_1(z) = \frac{j^2}{1-zj^2} + \frac{\lambda^2}{(1-2zj^2)^2} \tag{4.3.27}$$

The Green's function for a Wishart matrix is well known [65]

$$G_2(z) = \frac{1}{2vz}\left\{v(1-\alpha)+z-\sqrt{[(1-\alpha)v+z]^2-4vz}\right\} \tag{4.3.28}$$

where we recall that $\alpha = N^{\mathrm{s}}/N^{\mathrm{b}}$ and v, the variance, in our case is $v = k^2\sigma_{\mathrm{b}}^2/\sigma_{\mathrm{s}}^2$. The related R transform reads

$$R_2(z) = \frac{\alpha v}{1-vz} \tag{4.3.29}$$

The two above R-transforms now simply add to give the one for $\boldsymbol{M}$, $R_M(z) = R_1(z) + R_2(z)$. The result can be written as an implicit expression for the Green's function $G_M(z)$, given that from (4.3.22) one has generally $z(G) = 1/G + R(G)$

$$z = \frac{1}{G} + \frac{\alpha\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}}{1-\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}G} + \frac{j^2}{1-j^2G} + \frac{\lambda^2}{(1-2j^2G)^2} \tag{4.3.30}$$

We have abbreviated $G \equiv G_M$ on the r.h.s. here. Rearranging the above equation one sees that $G(z)$ is the solution of a fifth order polynomial equation. This can be found numerically, with the correct solution branch being determined from the asymptotic behaviour $G \approx 1/z$ for large z. Once $G(z)$ is in hand, $\rho(m)$ can be found using (4.3.24).

By a transformation of the spectrum of $\boldsymbol{M}$ we can characterize the spectrum of the posterior covariance matrix $\boldsymbol{C} = \sigma_{\mathrm{b}}^2\boldsymbol{M}^{-1/2}/2$ as well as the spectrum of relaxation rates as determined by the effective drift $\boldsymbol{K}^{\mathrm{bb|s}} = -\boldsymbol{M}^{1/2}$. The spectrum of $(-\boldsymbol{K}^{\mathrm{bb|s}})^{-1} = \boldsymbol{M}^{-1/2}$ then gives the distribution of relaxation times. As this matrix is proportional to $\boldsymbol{C}$, plots of $\rho(\tau)$ (figure 4.3) provide information also about the inference error as a function of α. The overall picture is that predictions become increasingly precise when the pool of observed data is expanded, i.e. α increases, while correlation times between posterior fluctuations decrease in proportion.

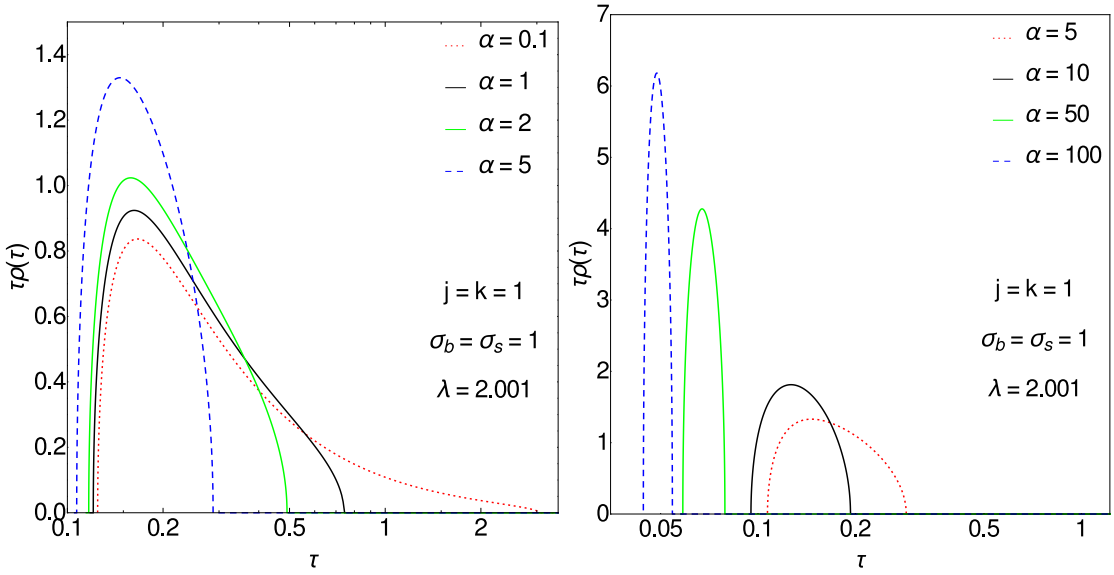

Figure 4.3: Spectral density $\rho(\ln\tau) = \tau\rho(\tau)$, of relaxation times τ, for different values of α. We plot $\rho(\ln\tau)$ to make the normalization of the densities more obvious. The spectra of posterior variances C, which define the inference error, are identical up to a horizontal shift as $C \propto \tau$. (Left) At small α the spectrum is broad, indicating that there is much variation in how different hidden state space directions are constrained by observations. For increasing α the spectrum becomes more peaked, and centred around decreasing τ or C: different directions become better determined, and more evenly, by observations, a trend more clearly visible on the right.

For qualitative analysis one can rewrite (4.3.30) in dimensionless variables $\tilde{z} = \sigma_s^2 z/(k^2\sigma_b^2)$ and $\tilde{G} = k^2\sigma_b^2 G/\sigma_s^2$ (in fact dimensionally $G \sim 1/z \sim k^2\sigma_b^2/\sigma_s^2$) as

$$\tilde{z} = \frac{1}{\tilde{G}} + \frac{\alpha}{1-\tilde{G}} + \frac{(\gamma p)^2}{1-(\gamma p)^2\tilde{G}} + \frac{p^2}{(1-2(\gamma p)^2\tilde{G})^2} \tag{4.3.31}$$

where $\gamma = j/\lambda$ and $p = \lambda/\sigma$, in analogy with chapter 3. This reduces the number of parameters and variables, from seven (α, j, k, λ, σ_s, σ_b, z) to four (p, γ, α, $\tilde{z}$). We recall that γ and $1/p$ measure the strength of hidden-hidden and hidden-observed couplings relative to the decay weight λ.

We have seen in figure 4.1 (left) that for $\gamma = 0$, i.e. in the absence of hidden-hidden interactions (see section 4.3.2) the spectrum consists of two separate pieces for $\alpha < 1$, while with such interactions present ($\gamma > 0$) the spectrum can be supported on a single interval. There must be a transition between these two cases at some value of γ that will depend on p and α - see figure 4.4 (left). Locating this transition numerically gives the results shown in figure 4.4 (right). The spectrum consists of a single piece *above* the line drawn in the (p, γ) plane. One sees that for large $p = \lambda/\sigma = \lambda\sigma_s/(\sigma_b k)$, i.e. weaker hidden-observed couplings, small values of $\gamma = j/\lambda$ and hence weak hidden-hidden interactions are sufficient to merge the two pieces of the spectrum.

Posterior correlations in Laplace space

From (4.2.9) and (4.3.21) we can obtain explicitly the posterior correlations in time: for $t > t'$,

$$\boldsymbol{C}(t-t') = \frac{\sigma_b^2}{2} e^{-\boldsymbol{M}^{\frac{1}{2}}(t-t')}\boldsymbol{M}^{-\frac{1}{2}} \tag{4.3.32}$$

In double-sided Laplace transform this becomes

$$\begin{aligned} \tilde{\boldsymbol{C}}(z) &= \frac{\sigma_b^2}{2}\left\{\int_{-\infty}^{0} e^{zs} e^{-\boldsymbol{M}^{\frac{1}{2}}s}\boldsymbol{M}^{-\frac{1}{2}}ds + \int_{0}^{+\infty} e^{-zs} e^{-\boldsymbol{M}^{\frac{1}{2}}s}\boldsymbol{M}^{-\frac{1}{2}}ds\right\} \\ &= \frac{\sigma_b^2}{2}\left\{\left[\boldsymbol{M}^{\frac{1}{2}}\left(z+\boldsymbol{M}^{\frac{1}{2}}\right)\right]^{-1} + \left[\boldsymbol{M}^{\frac{1}{2}}\left(-z+\boldsymbol{M}^{\frac{1}{2}}\right)\right]^{-1}\right\} \end{aligned} \tag{4.3.33}$$

We consider the trace, which at $t = t'$ gives the total posterior variance

$$\tilde{C}(z) = \frac{\sigma_b^2}{2}\,\mathrm{Tr}\left\{\left[\boldsymbol{M}^{\frac{1}{2}}\left(z+\boldsymbol{M}^{\frac{1}{2}}\right)\right]^{-1} + \left[\boldsymbol{M}^{\frac{1}{2}}\left(-z+\boldsymbol{M}^{\frac{1}{2}}\right)\right]^{-1}\right\} = -\sigma_b^2 G_M(z^2) \tag{4.3.34}$$

Such a relation is justified in light of the detailed balance condition satisfied by the posterior dynamics (4.2.6) with symmetric $\boldsymbol{J}$, as (4.3.34) can be seen as a re-formulation of the Fluctuation-Dissipation Theorem (FDT) [46]. According to FDT, one has

$$z\tilde{C}(z) = -\frac{\sigma_b^2}{2}[\tilde{R}(z) - \tilde{R}(-z)] \tag{4.3.35}$$

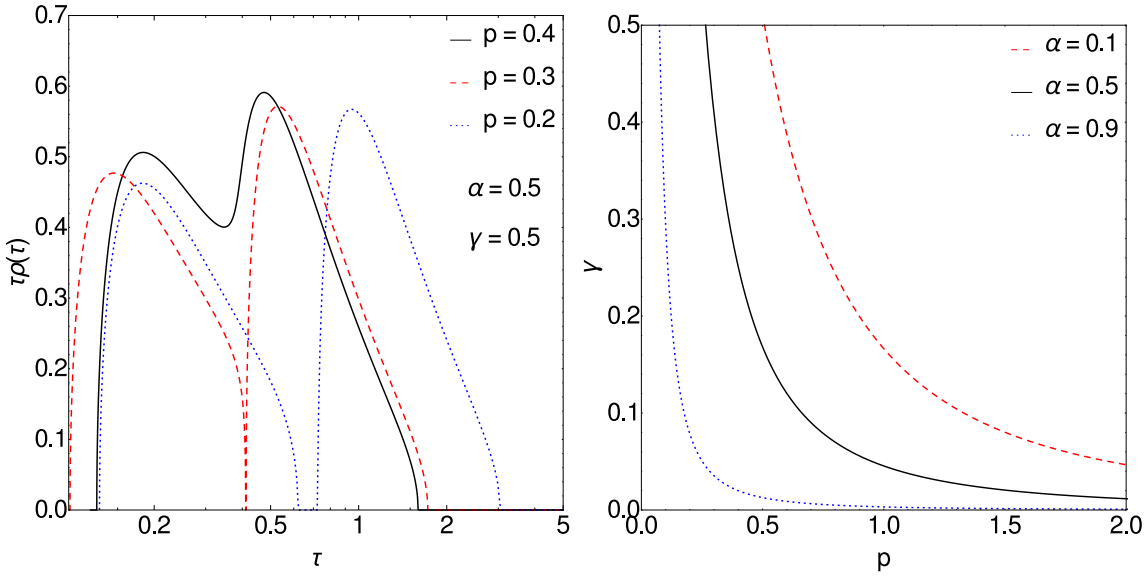

Figure 4.4: (Left) Spectral density $\rho(\ln\tau) = \tau\rho(\tau)$, at $\gamma = j/\lambda = 0.5$ (critical value for internal stability, with $j = 0.2$ and $\lambda = 0.4$) and $\alpha = 0.5$ for different values of p: the two pieces of the spectrum at $p = 0.2$ merge at $p = 0.3$, giving a spectrum supported on a single interval for $p > 0.3$. (Right) Curve in the p/γ space for which the two pieces of the spectrum merge when coming from low γ: the black line refers to $\alpha = 0.5$, the case shown on the left. The two-piece region near the origin shrinks (see curve for $\alpha = 0.9$) and vanishes for $\alpha \to 1$. The γ-axis is cut off at 0.5 as $\gamma > 0.5$ would be unphysical: we recall that the internal stability condition requires $\lambda > 2j$, which translates into $\gamma < 0.5$.

where $\tilde{R}(z)$ is the Laplace Transform of the response function for the posterior equation (4.2.6) evaluated in the thermodynamic limit, i.e.

$$\tilde{R}(z) = \mathrm{Tr}[z - \boldsymbol{K}^{\mathrm{bb|s}}]^{-1} \tag{4.3.36}$$

By FDT we have

$$\tilde{C}(z) = -\frac{\sigma_{\mathrm{b}}^2}{2z}\mathrm{Tr}\Big[(z - \boldsymbol{K}^{\mathrm{bb|s}})^{-1} - (-z - \boldsymbol{K}^{\mathrm{bb|s}})^{-1}\Big] = -\sigma_{\mathrm{b}}^2\mathrm{Tr}[z^2 - \boldsymbol{K}^{\mathrm{bb|s}\,2}]^{-1} = -\sigma_{\mathrm{b}}^2 G_M(z^2) \tag{4.3.37}$$

where $G_M(z^2)$ is the Green function of $\boldsymbol{K}^{\mathrm{bb|s}\,2} = \boldsymbol{M}$, exactly as in (4.3.34). We thus see that one can exploit the equilibrium property of the posterior dynamics (4.2.6) to derive information on $\tilde{C}(z)$ from the Green function of $\boldsymbol{K}^{\mathrm{bb|s}\,2}$. From (4.3.34), the Laplace transformed posterior correlation function has to satisfy the equation for $-\sigma_{\mathrm{b}}^2 G_M(z^2)$, giving

$$z^2 = -\frac{\sigma_{\mathrm{b}}^2}{\tilde{C}} + \frac{\alpha\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}}{1 + \frac{k^2}{\sigma_{\mathrm{s}}^2}\tilde{C}} + \frac{j^2}{1 + \frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}} + \frac{\lambda^2}{\left(1 + 2\frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}\right)^2} \tag{4.3.38}$$

where we have set $\tilde{C}(z) = \tilde{C}$. The Laplace Transform of the posterior covariance is then embedded as solution of this fifth order algebraic equation.

We incidentally notice that the equality (4.3.34) can be verified also for the purely self-interacting case (i.e. $j = 0$) as it trivially obeys the detailed balance condition. In this case, (4.3.38) becomes

$$z^2 - \lambda^2 = -\frac{\sigma_{\mathrm{b}}^2}{\tilde{C}} + \frac{\alpha\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}}{1 + \frac{k^2}{\sigma_{\mathrm{s}}^2}\tilde{C}} \tag{4.3.39}$$

This is the equation satisfied by $-\sigma_{\mathrm{b}}^2 G(z^2 - \lambda^2)$ of a Wishart matrix with $v = k^2\sigma_{\mathrm{b}}^2/\sigma_{\mathrm{s}}^2$. If we actually take $v = k^2\sigma_{\mathrm{b}}^2/\sigma_{\mathrm{s}}^2$, then we can recast expression (4.3.18) as

$$\begin{aligned}\tilde{C}(z) &= -\frac{1}{2v}\frac{\sigma_{\mathrm{b}}^2}{(z^2 - \lambda^2)}\left\{(1-\alpha)v + (z^2 - \lambda^2) + \sqrt{\left[(1-\alpha)v + (z^2 - \lambda^2)\right]^2 - 4(z^2 - \lambda^2)v}\right\} = \\ &= -\sigma_{\mathrm{b}}^2 G(z^2 - \lambda^2)\end{aligned} \tag{4.3.40}$$

i.e. $\tilde{C}(z)$ turns out to be, up to a sign and a factor σ_{b}^2, the resolvent of a Wishart matrix with variance v and calculated at $(z^2 - \lambda^2)$, see (4.3.28).

Interestingly, and similarly to (4.3.30) which determines the spectrum of $\boldsymbol{M}$, equation 4.3.38 does not become singular at $\lambda = 0$. This fact can be understood in the following way. If directions exist along which the hidden dynamics would grow exponentially without observations, then these always have a non-zero overlap with directions constrained by observed data. This is clear from the independent sampling of the two terms in $\boldsymbol{M}$, and explains how the posterior variance, the

uncertainty on the hidden dynamics, can stay finite even when the hidden dynamics without observations would diverge. We can draw the conclusion that the presence of observations stabilizes the dynamics regardless of their number (as long as one takes finite size observations) and α appears to play a role similar to the decay weight λ. Nevertheless, such a diverging hidden dynamics is an unphysical situation. We therefore continue to consider only parameter sets with $\lambda > \lambda_c$, the internal dynamical condition for a finite and well-defined marginal dynamics of the bulk.

Finally, by setting $z = i\omega$ one can evaluate the posterior power spectrum $\tilde{C}(i\omega)$. As done in chapter 3, the full expression for the conditional correlation can be recast as the product of an amplitude factor σ_s^2/k^2, accounting for its physical dimensions, and of a dimensionless function $C_{\alpha,p,\gamma}(\Omega)$

$$\tilde{C}(i\omega) = \frac{\sigma_s^2}{k^2} C_{\alpha,p,\gamma}(\Omega) \tag{4.3.41}$$

with $\Omega = \omega/\sigma$ a rescaled frequency. The prefactor shows that the entire power spectrum of the posterior variance or prediction uncertainty is directly proportional to the dynamical noise acting on the observed subnetwork σ_s^2 and inversely proportional to k^2, the strength with which it interacts with the bulk. As before (see (4.3.31)) one can find from (4.3.38) an equation for the dimensionless part C

$$-\Omega^2 = -\frac{1}{C} + \frac{\alpha}{1+C} + \frac{(\gamma p)^2}{1+(\gamma p)^2 C} + \frac{p^2}{(1+2(\gamma p)^2 C)^2} \tag{4.3.42}$$

where γ and p are defined as before.

One can verify that for $p = 0$ and $0 \leq \alpha \leq 1$, $C(0)$ has a divergence, implying also that the time integral of $\mathrm{Tr}\, \boldsymbol{C}(t - t')$ diverges. This comes physically from the fact that while a fraction α of hidden space directions have variances (and co-variances) of the expected order $\propto 1/k^2$, the others have variances that are independent of k and therefore much larger for large k.

A second region in the α, p, γ parameter space where $C(0)$ diverges is $\alpha \to 0$ and $\gamma \to \gamma_c = 1/2$. This is as expected: without observations, the hidden dynamics starts to diverge at $\lambda \to \lambda_c = 2j$, hence at $\gamma_c = 1/2$. We refer again to chapter 3 for further discussion of the behaviour in the vicinity of such critical points.

4.4 Thermodynamic Limit by Dynamical Functionals

So far we have studied the posterior variance and time-dependent covariance in settings where the dynamics of the entire network obeys detailed balance, and where the relevant Green's functions can be derived using RMT tools.

In the absence of detailed balance, dynamical functionals can be used as an alternative, within a statistical mechanics approach to inference (for a systematic discussion see [88, 107]).

We recall that the aim is to characterize a posterior path distribution, $P(\boldsymbol{X}^{\mathrm{b}}|\boldsymbol{X}^{\mathrm{s}})$, known to be Gaussian. The likelihood of the observed trajectory $P(\boldsymbol{X}^{\mathrm{s}})$ can be seen as a "partition function" Z that is obtained by summing $P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}})$ over all possible hidden paths $\boldsymbol{X}^{\mathrm{b}}$. From Z, one can define a free energy (density) to study macroscopic quantities such as mean and covariance of $P(\boldsymbol{X}^{\mathrm{b}}|\boldsymbol{X}^{\mathrm{s}})$. If the interactions are chosen randomly, they act as quenched disorder and the physically relevant quantity is the quenched average of the free energy,

$$f = -\lim_{N\to\infty} N^{-1} \langle \ln Z(\boldsymbol{J}, \boldsymbol{K}^{\mathrm{sb}}) \rangle_{\boldsymbol{J},\boldsymbol{K}^{\mathrm{sb}}} \tag{4.4.1}$$

where we have abbreviated $N^{\mathrm{b}} \equiv N$; Z is given by an average over hidden d.o.f. keeping the configuration of couplings to observations $\boldsymbol{K}^{\mathrm{sb}}$ fixed. The free energy $-N^{-1} \ln Z$ is self-averaging, i.e. its fluctuations around f for different realizations of the disorder vanish for $N \to \infty$. The same is true for the order parameters that arise in the calculation, which include the posterior variance, i.e. inference error.

Dynamical functionals appear in the above approach once we write the joint path probability $P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}})$ defined by the dynamics (4.2.1a) and (4.2.1b) in Onsager-Machlup form [18] as proportional to

$$\begin{aligned} P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}}) \propto \quad & \exp\left[-\frac{1}{2\sigma_{\mathrm{b}}^2} \int_0^T \left\| \partial_t \boldsymbol{x}^{\mathrm{b}} - \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{x}^{\mathrm{b}}(t) \right\|^2 dt \right] \\ & \cdot \exp\left[-\frac{1}{2\sigma_{\mathrm{s}}^2} \int_0^T \left\| \partial_t \boldsymbol{x}^{\mathrm{s}} - \boldsymbol{K}^{\mathrm{ss}} \boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t) \right\|^2 dt \right] \end{aligned} \tag{4.4.2}$$

with $\boldsymbol{K}^{\mathrm{bb}} = -\lambda \mathbb{1} + \boldsymbol{J}$. From the Gaussian form of this, the second order statistics of the posterior $P(\boldsymbol{X}^{\mathrm{b}}|\boldsymbol{X}^{\mathrm{s}})$ are independent of the value of the observed $\boldsymbol{X}^{\mathrm{s}}$. Hence to obtain the posterior variance it is sufficient to consider *zero observations*, i.e. $x_a(t) = 0$ for all a and t. All $\boldsymbol{x}^{\mathrm{b}}$ are then effectively deviations $\delta\boldsymbol{x}^{\mathrm{b}}$ from the posterior mean, though we will not write the δ explicitly to save space. The only remaining contribution from observations in (4.4.2) is in the couplings K_{aj} and the relevant partition function becomes

$$Z = \left\langle \exp\left[-\frac{1}{2\sigma_{\mathrm{s}}^2} \sum_{a=1}^{N^{\mathrm{s}}} \int_0^T \left(\sum_{j=1}^{N} K_{aj} x_j(t) \right)^2 dt \right] \right\rangle_{\boldsymbol{x}} \tag{4.4.3}$$

where $\boldsymbol{x} \equiv \boldsymbol{x}^{\mathrm{b}} = \{x_i\}_{i=1}^{N}$. The average is the marginalization over the hidden dynamics with the weight given by the second term in (4.4.2). This weight corresponds to the dynamics of the isolated hidden network, viz.

$$\partial_t x_i(t) = -\lambda x_i(t) + \sum_j J_{ij} x_j(t) + \xi_i(t) \tag{4.4.4}$$

with white noise $\langle \xi_i(t)\xi_j(t')\rangle = \sigma_b^2 \delta_{ij}\delta(t-t')$ as before.

4.4.1 Asymmetric hidden-hidden couplings

Annealed average

The average of $\ln Z$ over the quenched couplings $\boldsymbol{J}$ and $\boldsymbol{K}^{sb}$ would conventionally be performed by the replica method. However, for fully connected systems with quadratic interaction terms such as the one here, similar calculations [66, 92] indicate that the annealed calculation, which replaces $\langle \ln Z\rangle$ by $\ln\langle Z\rangle$, will give the exact result. We know in fact that order parameters which couple replicas α and β e.g. $\left\langle x^\alpha(t)x^\beta(t')\right\rangle = \frac{1}{N}\langle \sum_i m_i(t)m_i(t')\rangle_{\boldsymbol{J},\boldsymbol{K}} = 0$ for $\alpha \neq \beta$ with $m_i(t) = \langle x_i(t)\rangle_{\boldsymbol{x}}$ (assuming that we are in a steady state). This holds for linear, quadratic fully connected models, there is no "spin–glass" ordering: for the detailed outline of a similar calculation see [66, 92]. Hence, it is sufficient to perform the annealed average over the $\boldsymbol{J}$ and $\boldsymbol{K}^{sb}$ (i.e considering observations in equilibrium with other dynamical d.o.f.) and we calculate

$$f = -\lim_{N\to\infty} N^{-1} \ln\langle Z(\boldsymbol{J}, \boldsymbol{K}^{sb})\rangle_{\boldsymbol{J},\boldsymbol{K}^{sb}} \tag{4.4.5}$$

We shall again assume $\boldsymbol{J}$ and $\boldsymbol{K}^{sb}$ to have Gaussian-distributed elements with zero mean, but now consider the case where $\boldsymbol{J}$ is *asymmetric*, i.e. $\langle J_{ij}J_{ji}\rangle = 0$, thus breaking detailed balance. For the calculation we introduce

$$\chi_i(t) = \sum_{j=1}^{N} J_{ij}x_j(t) + \xi_i(t) \tag{4.4.6}$$

$$\phi_a(t) = \sum_{j=1}^{N} K_{aj}x_j(t) \tag{4.4.7}$$

With regards to the quenched disorder average these are two Gaussian fields, which become independent when conditioned on the x_i. Note that if $\boldsymbol{J}$ were not asymmetric, an additional memory/response function should be taken into account; for the sake of simplicity we show the baseline of the calculation for *asymmetric* couplings and will just provide the result for general symmetry in section 4.4.2. Defining as before amplitudes j and k so that $\langle J_{ij}^2\rangle = j^2/N$ and $\langle K_{aj}^2\rangle = k^2/N$, we have

$$\langle \chi_i(t)\chi_i(t')\rangle_{\boldsymbol{J}} = \sigma_b^2\delta(t-t') + \frac{j^2}{N}\sum_{j=1}^{N} x_j(t)x_j(t') = \sigma_b^2\delta(t-t') + j^2C(t,t') \tag{4.4.8}$$

$$\langle \phi_a(t)\phi_b(t')\rangle_{\boldsymbol{J}} = k^2C(t,t')\delta_{ab} \tag{4.4.9}$$

where we have introduced the order parameter

$$C(t,t') \doteq \frac{1}{N}\sum_{j=1}^{N} x_j(t)x_j(t') \tag{4.4.10}$$

Hence, we will calculate

$$Z_{\rm ann} = \left\langle \exp\left[\frac{1}{2\sigma_{\rm s}^2}\sum_{a=1}^{N^{\rm s}}\int_0^T \phi_a^2(t)dt\right]\right\rangle_{\boldsymbol{\phi},\boldsymbol{x}} \tag{4.4.11}$$

where now the process has an effective prior dynamics given by

$$\partial_t x_i(t) = -\lambda x_i(t) + \chi_i(t) \tag{4.4.12}$$

Here $\boldsymbol{\phi} = \{\phi_a\}_{a=1}^{N^{\rm s}}$ and $\boldsymbol{\chi} = \{\chi_i\}_{i=1}^{N}$ are still coupled to $\boldsymbol{x}$ because of the covariances $C(t,t')$.

Decoupling the degrees of freedom

To decouple the degrees of freedom we constrain the value of the order parameter function $C(t,t')$. Formally this means writing $Z_{\rm ann}$ as an integral of $\exp(N\Xi[C])$ over all possible values of $C(t,t')$, where

$$\begin{aligned}\Xi[C] &= \frac{1}{N}\ln\left\langle \exp\left\{-\frac{1}{2\sigma_{\rm s}^2}\sum_{a=1}^{N^{\rm s}}\int_0^T \phi_a^2(t)dt\right\}\prod_{t,t'}\delta\left(NC(t,t') - \sum_{i=1}^{N} x_i(t)x_i(t')\right)\right\rangle_{\boldsymbol{\phi},\boldsymbol{x}} \\ &\equiv \Xi_1[C] + \Xi_2[C]\end{aligned} \tag{4.4.13}$$

with

$$\Xi_1[C] = \frac{1}{N}\ln\left\langle \prod_{t,t'}\delta\left(NC(t,t') - \sum_{i=1}^{N} x_i(t)x_i(t')\right)\right\rangle_{\boldsymbol{x}} \tag{4.4.14}$$

$$\Xi_2[C] = \frac{N^{\rm s}}{N}\ln\left\langle \exp\left\{-\frac{1}{2\sigma_{\rm s}^2}\int_0^T \phi^2(t)dt\right\}\right\rangle_{\phi} \tag{4.4.15}$$

In equation (4.4.15) the decoupling has allowed us to drop the index a and consider a representative ϕ. The first equation (4.4.14) is dealt with by introducing an order parameter to $C(t,t')$. This means that for $N \to \infty$, we replace the "hard" δ constraints by an extra Gaussian term yielding a new effective measure over independent $x_i(t)$, which is adjusted such that $\langle x_i(t)x_i(t')\rangle_e = C(t,t')$ (here e denotes the effective "posterior" average). Equivalently one can write δ-function constraints in Fourier representation and evaluate $\exp(N\Xi[C])$ using a saddle point method. Either way one has

$$\Xi_1 = \frac{1}{2}\int_0^T dt\int_0^T dt'\; D(t,t')C(t,t') + \ln\left\langle \exp\left\{-\frac{1}{2}\int_0^T dt\int_0^T dt'\; D(t,t')x(t)x(t')\right\}\right\rangle_x \tag{4.4.16}$$

This path integral is now also for a single representative coordinate x (we dropped index i as all the i become independent and equivalent). Extremization over $D(t,t')$ is understood in (4.4.16), and similarly one needs to extremize over $C(t,t')$ in evaluating the resulting $Z_{\rm ann}$.

Evaluating the order parameters

As before we focus on the steady state of the system for $t \to \infty$. The order parameters then depend on time differences only and the path integrals can be evaluated using Fourier or Laplace modes $\tilde{x}(z)$. These decouple into independent Gaussians and we get from (4.4.8), (4.4.9) and (4.4.12) that

$$\tilde{C}_0(z) \doteq \left\langle |\tilde{x}(z)|^2 \right\rangle_{\tilde{x}} = \frac{j^2 \tilde{C}(z) + \sigma_{\rm b}^2}{-z^2 + \lambda^2} \tag{4.4.17}$$

$$\left\langle |\tilde{\phi}(z)|^2 \right\rangle_{\tilde{\phi}} = k^2 \tilde{C}(z) \tag{4.4.18}$$

$\tilde{C}_0(z)$ is the covariance of the prior effective dynamics while $\tilde{C}(z)$ relates to the posterior dynamics that includes the conditioning on observations. Carrying out the prior average, the second term in (4.4.16) becomes

$$\ln \left\langle \exp\left\{ -\frac{1}{2} \int_0^T dt \int_0^T dt' \, D(t,t') x(t) x(t') \right\} \right\rangle_x = \tag{4.4.19}$$

$$= \int dz \ln \int_{-\infty}^{\infty} \frac{d\tilde{x}(z)}{\sqrt{2\pi \tilde{C}_0(z)}} \exp\left\{ -\frac{|\tilde{x}(z)|^2}{2} \left(\tilde{C}_0(z)^{-1} + \tilde{D}(z) \right) \right\} =$$

$$= \int dz \ln \left(\frac{1}{\sqrt{2\pi \tilde{C}_0(z)}} \sqrt{\frac{2\pi}{\tilde{C}_0(z)^{-1} + \tilde{D}(z)}} \right) = -\frac{1}{2} \int dz \ln\left(1 + \tilde{C}_0(z) \tilde{D}(z) \right) \tag{4.4.20}$$

The second line gives explicitly the average over the Gaussian distribution in Laplace domain over the "prior" trajectories (4.4.12) with covariance $\tilde{C}_0(z)$. In a similar way, we have for Ξ_2, from (4.4.15)

$$\Xi_2[C] = \left\langle \exp\left\{ -\frac{1}{2\sigma_{\rm s}^2} \int_0^T \phi^2(t) dt \right\} \right\rangle_\phi = \int dz \ln \int_{-\infty}^{\infty} \frac{d\tilde{\phi}(z)}{\sqrt{2\pi k^2 \tilde{C}(z)}} \exp\left\{ -\frac{|\tilde{\phi}(z)|^2}{2} \left(\frac{1}{k^2 \tilde{C}(z)} + \frac{1}{\sigma_{\rm s}^2} \right) \right\} =$$

$$= -\frac{1}{2} \int dz \ln\left(1 + \frac{k^2}{\sigma_{\rm s}^2} \tilde{C}(z) \right) \tag{4.4.21}$$

In the second line we have performed the average over the Gaussian field ϕ that has covariance (4.4.9). Hence, finally, by substituting (4.4.19) into (4.4.16) and from (4.4.21) we get

$$\Xi = \frac{1}{2} \int dz \left[\tilde{D}(z) \tilde{C}(z) - \ln\left(1 + \tilde{C}_0(z) \tilde{D}(z) \right) \right] - \frac{\alpha}{2} \int dz \ln\left(1 + \frac{k^2}{\sigma_{\rm s}^2} \tilde{C}(z) \right) \tag{4.4.22}$$

where $\alpha = N^{\rm s}/N$ as before. The order parameter equations $\partial \Xi / \partial \tilde{C}(z) = 0$ and $\partial \Xi / \partial \tilde{D}(z) = 0$ result as

$$\tilde{D}(z) = \frac{\alpha k^2}{\sigma_{\rm s}^2 + k^2 \tilde{C}(z)} + \frac{\tilde{D}(z)}{1 + \tilde{C}_0(z) \tilde{D}(z)} \frac{j^2}{-z^2 + \lambda^2} \tag{4.4.23}$$

$$\frac{\tilde{C}(z)}{\tilde{C}_0(z)} + \tilde{D}(z) \tilde{C}(z) = 1 \tag{4.4.24}$$

We shall notice that, in absence of observations (i.e. $\alpha = 0$), the normalization of the probability distribution over hidden variables implies $Z_{\rm ann} = 1$, which would translate into $\tilde{D}(z) \equiv 0 \ \forall z$.

In addition, the distinction between the prior $\tilde{C}_0(z)$ and the posterior $\tilde{C}(z)$ would not hold true any more, thus $\tilde{C}_0(z) \equiv \tilde{C}(z)\ \forall z$: equations (4.4.23) and (4.4.24) are actually consistent with this particular case. Combining equations (4.4.23) and (4.4.24) and using (4.4.17) gives a closed algebraic equation for $\tilde{C}(z)$

$$z^2 = \left[-\frac{\sigma_b^2}{\tilde{C}} + \frac{\alpha \frac{k^2\sigma_b^2}{\sigma_s^2}}{1 + \frac{k^2}{\sigma_s^2}\tilde{C}} \right] \left(1 + \frac{j^2}{\sigma_b^2}\tilde{C}\right)^2 + j^2\left(1 + \frac{j^2}{\sigma_b^2}\tilde{C}\right) + \lambda^2 \tag{4.4.25}$$

with the abbreviation $\tilde{C}(z) = \tilde{C}$. This is the analogue of (4.3.38) for the non-equilibrium case of asymmetric couplings $\boldsymbol{J}$ and our final result for this section.

4.4.2 Generalization to arbitrary interaction symmetry

The above approach based on dynamical functionals can be extended to the case of hidden-hidden interactions of arbitrary degree of symmetry, defined by $\langle J_{ij}J_{ji}\rangle = \eta j^2/N$. Asymmetric couplings (section 4.4.1) correspond to $\eta = 0$ while $\eta = 1$ gives symmetric $\boldsymbol{J}$ (section 4.3.3). The $N \to \infty$ spectrum of these RMs is described by the Girko's elliptic law [43], i.e. it is uniform in an ellipse in the complex plane with semi-axes $j(1+\eta)$ and $j(1-\eta)$, respectively, along the real and imaginary directions. The critical value of λ is defined as the largest positive real eigenvalue of $\boldsymbol{J}$, thus for this kind of hidden interactions one gets $\lambda_c = j(1+\eta)$. The main change is that the nonzero correlation $\langle J_{ij}J_{ji}\rangle$ causes the effective prior dynamics to contain a response term where each $x_i(t)$ reacts to its values $x_i(t')$ in the past. This feedback of previous times, modulated by a response function, should be included via an integral term (see e.g. [50]); the strength of such a coupling to the past is controlled by the intensity ηj^2

$$\partial_t x_i(t) = -\lambda x_i(t) + \eta j^2 \int_0^t dt' R(t,t') x_i(t') + \chi_i(t) \tag{4.4.26}$$

where $R(t,t')$ is the average macroscopic response of the hidden system at time t to a perturbation at t'. Following the same logical thread as in section 4.4.1 one can show that the dynamics can be decoupled by introducing order parameter functions $R(t,t')$ and $C(t,t')$; in particular, given the new prior dynamics (4.4.26), the expression (4.4.17) of $\tilde{C}_0(z)$ will be modified

$$\tilde{C}_0(z) \doteq \left\langle |\tilde{x}(z)|^2 \right\rangle_{\tilde{x}} = \frac{j^2\tilde{C}(z) + \sigma_b^2}{[-z + \lambda - \eta j^2 \tilde{R}(-z)][z + \lambda - \eta j^2 \tilde{R}(z)]} \tag{4.4.27}$$

and, proceeding as previously, the final equations read

$$\tilde{D}(z) = \frac{\alpha k^2}{\sigma_s^2 + k^2\tilde{C}(z)} + \frac{\tilde{D}(z)}{1 + \tilde{C}_0(z)\tilde{D}(z)} \frac{j^2}{[-z + \lambda - \eta j^2 \tilde{R}(-z)][z + \lambda - \eta j^2 \tilde{R}(z)]} \tag{4.4.28}$$

$$\frac{\tilde{C}(z)}{\tilde{C}_0(z)} + \tilde{D}(z)\tilde{C}(z) = 1 \tag{4.4.29}$$

We stress that also the response $\tilde{R}(z)$ must be considered as an order parameter, thus we have to minimize $\Xi[C, R, D]$ also w.r.t. it. This condition leads one to

$$\frac{\partial \Xi}{\partial \tilde{R}(z)} = -\frac{1}{2}\frac{\tilde{D}(z)}{(1+\tilde{C}_0(z)\tilde{D}(z))}\frac{\partial \tilde{C}_0(z)}{\partial \tilde{R}(z)} = 0 \tag{4.4.30}$$

From (4.4.29) we can write $\tilde{D}(z)/(1+\tilde{C}_0(z)\tilde{D}(z)) = (1-\tilde{C}(z)/\tilde{C}_0(z))/\tilde{C}_0(z)$; also we know that the response of the dynamics (4.4.26) obeys $\tilde{R}(z) = (z+\lambda-\eta j^2\tilde{R}(z))^{-1}$ (see chapter 2 for a derivation). By using these results and by plugging the definition (4.4.27) into (4.4.30) one obtains

$$\tilde{C}(z)[-z+\lambda-\eta j^2\tilde{R}(-z)] - \tilde{R}(z)[j^2\tilde{C}(z)+\sigma_{\mathrm{b}}^2] = 0 \tag{4.4.31}$$

If we set $\tilde{D}(z) = -\Big(-\frac{\alpha k^2}{\sigma_{\mathrm{s}}^2+k^2\tilde{C}(z)} + j^2\tilde{B}(z)\Big)$ we see that (4.4.28) and (4.4.29) can be recast as

$$\begin{aligned}\tilde{C}(z)\Big[&\Big(-z+\lambda-\eta j^2\tilde{R}(-z)\Big)\Big(j^2\tilde{C}(z)+\sigma_{\mathrm{b}}^2\Big)^{-1}\Big(z+\lambda-\eta j^2\tilde{R}(z)\Big)\\ &+\frac{\alpha k^2}{\sigma_{\mathrm{s}}^2+k^2\tilde{C}(z)} - j^2\tilde{B}(z)\Big] = 1\end{aligned} \tag{4.4.32}$$

$$\begin{aligned}\tilde{B}(z)\Big[&\Big(-z+\lambda-\eta j^2\tilde{R}(-z)\Big)\Big(-\frac{\alpha k^2}{\sigma_{\mathrm{s}}^2+k^2\tilde{C}(z)} + j^2\tilde{B}(z)\Big)^{-1}\Big(z+\lambda-\eta j^2\tilde{R}(z)\Big)\\ &-j^2\tilde{C}(z)-\sigma_{\mathrm{b}}^2\Big] = 1\end{aligned} \tag{4.4.33}$$

This system of equations, together with the $\tilde{R}(z)$ equation (4.4.31), is the same as the one obtained by the Extended Plefka Expansion in chapter 3: in absence of observations, i.e. $\alpha = 0$, it is shown there that $\tilde{B}(z) \equiv 0$, which is consistent with a vanishing $\tilde{D}(z)$. As further check, one can also verify that algebraic equation (4.3.38) can be recovered from (4.4.28) and (4.4.29) by setting $\eta = 1$.

By solving (4.4.31) (taken for z and $-z$) in terms of $\tilde{C}(z)$ we obtain

$$\tilde{R}(z) = \tilde{C}(z)\Big[\frac{\lambda}{\sigma_{\mathrm{b}}^2+j^2(1+\eta)\tilde{C}(z)} - \frac{z}{j^2(1-\eta)\tilde{C}(z)+\sigma_{\mathrm{b}}^2}\Big] \tag{4.4.34}$$

thus

$$[-z+\lambda-\eta j^2\tilde{R}(-z)][z+\lambda-\eta j^2\tilde{R}(z)] = (j^2\tilde{C}(z)+\sigma_{\mathrm{b}}^2)^2\Big[\Big(\frac{\lambda}{\sigma_{\mathrm{b}}^2+j^2(1+\eta)\tilde{C}(z)}\Big)^2 - \Big(\frac{z}{j^2(1-\eta)\tilde{C}(z)+\sigma_{\mathrm{b}}^2}\Big)^2\Big] \tag{4.4.35}$$

One can express also $\tilde{D}(z)$ as a function of $\tilde{C}(z)$ from (4.4.28) using (4.4.35), so that by substitution into (4.4.29) the final result is again a closed algebraic equation for $\tilde{C}(z)$

$$z^2 = \left[-\frac{\sigma_{\mathrm{b}}^2}{\tilde{C}} + \frac{\alpha\frac{k^2\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{s}}^2}}{1+\frac{k^2}{\sigma_{\mathrm{s}}^2}\tilde{C}} + \frac{j^2}{1+\frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}} + \frac{\lambda^2}{\Big(1+(1+\eta)\frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}\Big)^2}\right]\Big(1+(1-\eta)\frac{j^2}{\sigma_{\mathrm{b}}^2}\tilde{C}\Big)^2 \tag{4.4.36}$$

For $\eta = 1$ and $\eta = 0$ this leads back to (4.3.38) and (4.4.25), respectively, as it should.

The result (4.4.36) characterizes the average case posterior variance – and hence inference error – for our partially observed network dynamics. Remarkably, it does so across an entire range of non-equilibrium settings parameterized by η. Equation (4.4.36) is derived within the annealed approximation but as discussed above this should be exact here so that our result acts as a baseline for the assessment of other approximations. One such approximation, the Extended Plefka Expansion (chapter 3) give precisely (4.4.36), demonstrating that this approximate scheme is also exact (in the large system limit studied here). As a consequence, we can refer to chapter 3 for a systematic analysis of inference errors and posterior relaxation times as they result from (4.4.36). This analysis can be organized around critical regions in the parameter space of α, γ and p. We recall that there are two such regions for general symmetry parameter η, similarly to what we found in section 4.3.3: one is defined by $p \to 0$ for $0 \leq \alpha \leq 1$, the second by $\alpha \to 0$ and $\gamma \to \gamma_c = 1/(1+\eta)$. Let us finally stress that such a configuration of singularities in the α, γ, p space, entailing γ_c as critical point only with no observations ($\alpha \to 0$), suggests also in this case that the conditional dynamics can be stable even when the marginal hidden one exhibits divergent trajectories. Indeed the argument put forward in section 4.3.3 justifying the stability of conditional quantities does not rely on the symmetry of $\boldsymbol{K}^{\mathrm{bb}}$, i.e. it is applicable, as expected, to the case of generic symmetry η.

4.5 Discussion and Conclusion

We have considered in this chapter a linear stochastic dynamics in a large network of continuous degrees of freedom, where given a time trajectory of the nodes in some observable part of the network the task is to infer the trajectory of the hidden nodes. By varying interaction symmetry we were able to study both equilibrium and non-equilibrium settings, thus creating a paradigmatic example of inference from temporal data. Given the increasing availability of large scale temporal data sets such problems are becoming prevalent in e.g. biology, where interpretation of data and prediction are highly challenging when observations only partially characterize a system.

Our main goal was to explore the average case inference error. To ensure analytical tractability we focussed on stationary dynamics on large networks. More precisely it is the variance of hidden state estimates that becomes stationary in time; mean predictions for the hidden states have to depend on time in our dynamical context.

The large network assumption is realistic in many situations, e.g. for metabolic or neural networks that can be composed of thousands of interacting elements (chemical species, neurons etc).

We deployed two different methods of analysis. For the first, the starting point (section 4.2) is a

Lyapunov-type equation for the posterior variance matrix $\boldsymbol{C}$, where an effective drift matrix $\boldsymbol{K}^{\mathrm{bb|s}}$ captures the effect of the observations. In this picture, the inference problem can be recast in terms of an "effective" kinetics, i.e. where rates are renormalized by conditioning on observations.

In section 4.3 we derived average case performance results by appeal to RMT. This is possible because the Lyapunov equation can be solved in the case of self-interacting hidden variables (section 4.3.2) or more generally, symmetric hidden-hidden couplings (section 4.3.3), corresponding to equilibrium dynamics. With suitable assumptions of couplings being Gaussian and long-range, and taking the thermodynamic limit of large networks, we then used free probability methods to derive the Green's functions and then the spectra of $\boldsymbol{C}$ and $\boldsymbol{K}^{\mathrm{bb|s}}$, which are closely linked.

For the opposite case of *asymmetric* hidden-hidden couplings, where the dynamics is non-equilibrium, we presented in section 4.4.1 a calculation based on dynamical functionals. This leads to an algebraic equation for the stationary posterior variance (in Laplace space). We sketched how the approach can be extended to the analysis of non-equilibrium stationary regimes arising from couplings of *generic* symmetry (section 4.4.2).

We focussed on the inference error as an average macroscopic quantity. For large networks this is independent of the specific realization of the microscopic (Gaussian) interactions, but does depend on structural parameters such as overall interaction strengths as well as α, the ratio between the number of hidden and observed nodes. Predictions on such structural dependences of macroscopic properties should be testable in practice and may give information on microscopic features such as the degree of interaction symmetry.

The RMT approach to our problem has the benefit that it gives information on spectral densities, including the spectrum of relaxation times in the posterior dynamics. This then allowed us to compare different definitions of a characteristic posterior relaxation time, such as slowest mode and average time (section 4.3.2). The spectral shapes proved revealing: when there are few observations (small α), the spectrum can be split into two parts corresponding to constrained and unconstrained directions (section 4.3.3), but this distinction is then lost as hidden nodes interact more strongly.

In sections 4.4.1 and 4.4.2, we found as well algebraic equations for the average response function, whose singularities are a feature to explore for the spectrum characterization. Nevertheless, as highlighted in the development of the Extended Plefka Expansion (chapter 2), the correlation function seems to embed less information on the structure of the spectrum support than the response function (e.g. in chapter 2 the responses "see" the edge of the ellipses in the sense that they become singular for those values). The expression of the response given by the analytic

continuation from infinite z is valid only outside the spectrum, as also the electrostatic analogy shows [45]. In and out the spectrum, the physical solution of the average response is given by different branches, which meet on the edge of the spectrum. The regularization technique would allow one to follow smoothly this change, thus to go from outside to inside in a controlled way [44]: one could think of applying this technique to the responses obtained in this chapter to explore the support of $\boldsymbol{K}^{\mathrm{bb|s}}$ spectrum. One open question for the inference setting we have considered is then to answer the question of the spectral density of relaxation times and its support in the *non-equilibrium* case $\eta < 1$. For example, does our result (4.4.36) for generic η still have a free probability interpretation? Generalizing the derivation of the equilibrium ($\eta = 1$) result (4.3.38) to $\eta < 1$ appears non-trivial. The derivation in the latter case relied essentially on the symmetry of couplings and the FDT, thus it is not obvious at all that a generalization for any symmetry could actually be implementable. One might consider assuming that the equilibrium relation $\tilde{C}(z) = -\sigma_{\mathrm{b}}^2 \tilde{G}(z^2)$ continues to hold effectively and analyze the spectrum corresponding to the Green's function $\tilde{G}(z)$. More generally, it would be interesting to see whether the RMT-based analysis could be expanded, in such a way to yield an exact characterization of the inference error also for non-equilibrium hidden dynamics.

There are a number of avenues for further work, as the setting we have begun to study is still rather new in the statistical physics community [37, 57–59]. An obvious extension would be to sparse networks, where for static analyses statistical mechanics has been successfully deployed [108, 109]. The sparse case would be worth developing because of its relevance to applications such as gene expression networks [85], although it might raise nontrivial mathematical difficulties.

Variants of the dynamics could also be considered, for example, by adding non-linearities that can be treated perturbatively. One could also extend to measurements of the trajectory of the observable nodes that would be available at a regular or irregular grid of time points only rather than along the entire time interval considered; or to measurements which are noisy rather than just incomplete as in our case [82, 110]. So far we have in fact considered the case of incomplete but not corrupted state information, i.e. not all the state variables are measured but, if they are available, they are precisely known.

Finally, as in chapter 3, we have concentrated on the *forward* problem of predicting hidden states given known interactions. This is relevant also for inverse problems such as learning the couplings from dynamical data, where typically a forward problem has to be solved at every iteration (e.g. in Expectation Propagation [54, 55]). Learning which couplings are non-zero is effectively a network reconstruction problem, with potential applications to signalling pathways and gene

expression data. Algorithmic advances have already been achieved by adapting equilibrium statistical physics tools [85, 111] to learning of regulatory networks from steady state data. In our case, modelling data as explicitly dynamical rather than as uncorrelated snapshots is expected to lead to performance improvements in inference and learning.

Appendix E

Kalman filter and smoother

E.1 Kalman filter and smoother

In this appendix we derive the results (4.2.5)-(4.2.9) in the main text, using a reduction of our inference problem to a linear Gaussian state space model, to which standard Kalman filter techniques [16] can then be applied.

Let us consider a time discretized version of our dynamics (4.2.1a) and (4.2.1b), with elementary time step Δ,

$$\boldsymbol{x}^{\mathrm{b}}(t) - \boldsymbol{x}^{\mathrm{b}}(t-\Delta) = \Delta \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t-\Delta) + \Delta \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{x}^{\mathrm{b}}(t-\Delta) + \Delta \bar{\boldsymbol{\xi}}^{\mathrm{b}}(t-\Delta) \tag{E.1.1a}$$

$$\boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{x}^{\mathrm{s}}(t-\Delta) = \Delta \boldsymbol{K}^{\mathrm{ss}} \boldsymbol{x}^{\mathrm{s}}(t-\Delta) + \Delta \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t-\Delta) + \Delta \bar{\boldsymbol{\xi}}^{\mathrm{s}}(t-\Delta) \tag{E.1.1b}$$

where the white noises $\bar{\boldsymbol{\xi}}^{\mathrm{s}}$ and $\bar{\boldsymbol{\xi}}^{\mathrm{b}}$ are averages of the continuous time noise over the time interval Δ with covariance

$$\langle \bar{\boldsymbol{\xi}}^{\mathrm{s}}(t) \bar{\boldsymbol{\xi}}^{\mathrm{s}\,T}(t') \rangle = \Delta^{-1} \boldsymbol{\Sigma}^{\mathrm{ss}} \delta_{tt'} \tag{E.1.2}$$

and similarly for $\bar{\boldsymbol{\xi}}^{\mathrm{b}}$.

The above dynamics is Markovian, with transition probabilities

$$P(\boldsymbol{x}^{\mathrm{b}}(t) | \boldsymbol{x}^{\mathrm{b}}(t-\Delta), \boldsymbol{x}^{\mathrm{s}}(t-\Delta)) = \tag{E.1.3}$$
$$\mathcal{N}(\boldsymbol{x}^{\mathrm{b}}(t) | (\mathbb{1} + \Delta \boldsymbol{K}^{\mathrm{bb}}) \boldsymbol{x}^{\mathrm{b}}(t-\Delta) + \Delta \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t-\Delta), \Delta \boldsymbol{\Sigma}^{\mathrm{bb}})$$

$$P(\boldsymbol{x}^{\mathrm{s}}(t+\Delta) | \boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) = \tag{E.1.4}$$
$$\mathcal{N}(\boldsymbol{x}^{\mathrm{s}}(t+\Delta) | (\mathbb{1} + \Delta \boldsymbol{K}^{\mathrm{ss}}) \boldsymbol{x}^{\mathrm{s}}(t) + \Delta \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t), \Delta \boldsymbol{\Sigma}^{\mathrm{ss}})$$

and we are interested in the posterior probability $P(\boldsymbol{X}^{\mathrm{b}} | \boldsymbol{X}^{\mathrm{s}})$ of a time trajectory $\boldsymbol{X}^{\mathrm{b}}$ of hidden variables given a trajectory $\boldsymbol{X}^{\mathrm{s}}$ of observed variables.

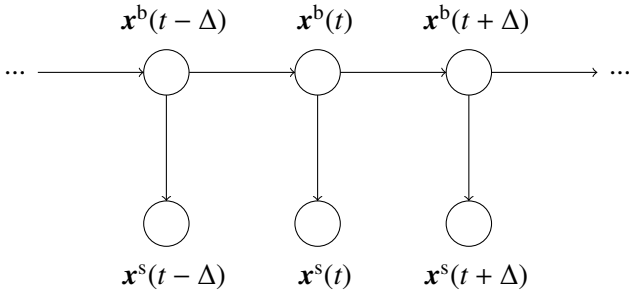

Figure E.1: Illustration of a linear-Gaussian state space model.

To bring this inference problem into a standard form, we exploit the fact that the joint distribution $P(\boldsymbol{X}^b, \boldsymbol{X}^s)$ is Gaussian, and hence so is the posterior $P(\boldsymbol{X}^b|\boldsymbol{X}^s)$. From general properties of Gaussian conditioning, the second order statistics of the posterior are then *independent* of the specific observed trajectory $\boldsymbol{X}^s$. We can therefore choose the most convenient $\boldsymbol{X}^s$ to find the second order statistics, which is the identically zero trajectory. The second order statistics we find then determine the inference error, which is the (normalized) trace of the covariance matrix of $\boldsymbol{x}^b(t)$.

For zero observations, the transition probabilities (E.1.3), (E.1.4) simplify to

$$P(\boldsymbol{x}^b(t)|\boldsymbol{x}^b(t-\Delta)) = \mathcal{N}(\boldsymbol{x}^b(t)|(\mathbb{1} + \Delta \boldsymbol{K}^{bb})\boldsymbol{x}^b(t-\Delta), \Delta\boldsymbol{\Sigma}^{bb}) \tag{E.1.5}$$

$$P(\boldsymbol{x}^s(t+\Delta) = 0|\boldsymbol{x}^b(t)) = \mathcal{N}(\boldsymbol{x}^s(t+\Delta) = 0|\Delta \boldsymbol{K}^{sb}\boldsymbol{x}^b(t), \Delta\boldsymbol{\Sigma}^{ss}) \tag{E.1.6}$$

These now have the conventional form of a linear-Gaussian state space model [16], where (E.1.5) specifies the dynamics of the hidden state $\boldsymbol{x}^b$ while (E.1.6) defines the "emission probability" at time t, with $\boldsymbol{x}^s(t+\Delta)$ taking the role of the emitted signal or observation. To conform with standard notation, we will shift the time index on $\boldsymbol{x}^s(t+\Delta)$ to $\boldsymbol{x}^s(t)$ for the rest of this discussion; see figure E.1. Note that while we are dealing with real-valued states and emissions here, the probabilistic "graphical model" of figure E.1 could also capture cases, e.g. *Hidden Markov Models* (HMMs) with the hidden states are discrete [16]. The chain structure of figure E.1 means that posterior probabilities can be computed efficiently by message passing methods, denoted *Forward-Backward* algorithm in the context of HMMs [112] and *Kalman Filter* [63] [1].

The forward propagation computes forward messages $\hat{\alpha}_t$ that absorb the effect of previous observations (the past), while the backward propagation $\hat{\beta}_t$ accounts for observations from the future. Formally the messages can be defined as

$$\hat{\alpha}(\boldsymbol{x}^b(t)) = P(\boldsymbol{x}^b(t)|\boldsymbol{x}^s(\Delta), ..., \boldsymbol{x}^s(t)) = \hat{\alpha}_t, \tag{E.1.7}$$

[1] Rigorously only the recursive computation of forward messages should be referred to as Kalman filter, while equations of backward messages are known as Kalman smoothers. The original paper by Kalman [63] addresses only the problem of filtering, i.e. recovery of hidden dynamics at time t using measurements up to t.

$$\hat{\beta}(\boldsymbol{x}^{\mathrm{b}}(t)) \quad = \quad \frac{P(\boldsymbol{x}^{\mathrm{s}}(t+\Delta),...,\boldsymbol{x}^{\mathrm{s}}(T)|\boldsymbol{x}^{\mathrm{b}}(t))}{P(\boldsymbol{x}^{\mathrm{s}}(t+\Delta),...,\boldsymbol{x}^{\mathrm{s}}(T)|\boldsymbol{x}^{\mathrm{s}}(\Delta),...,\boldsymbol{x}^{\mathrm{s}}(t))} = \hat{\beta}_t \tag{E.1.8}$$

Once $\hat{\alpha}_t$ and $\hat{\beta}_t$ have been computed, the desired posterior probability is simply

$$\gamma_t = \hat{\alpha}_t\hat{\beta}_t = \frac{P(\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{X}^{\mathrm{s}})}{P(\boldsymbol{X}^{\mathrm{s}})} = P(\boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{X}^{\mathrm{s}}) \tag{E.1.9}$$

The forward propagation for continuous variables reads

$$\hat{\alpha}_t \quad \propto \quad P(\boldsymbol{x}^{\mathrm{s}}(t)|\boldsymbol{x}^{\mathrm{b}}(t)) \int d\boldsymbol{x}^{\mathrm{b}}(t-\Delta) P(\boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{x}^{\mathrm{b}}(t-\Delta))\hat{\alpha}_{t-\Delta} \tag{E.1.10}$$

For a linear-Gaussian model, emission and transition probabilities appearing in (E.1.10) are Gaussian; in this case, they are given by (E.1.5) and (E.1.6).

By closure properties of the Gaussian family, all distributions involved, $\hat{\alpha}_t$, $\hat{\beta}_t$ and γ_t, are also Gaussian and we denote in particular

$$\hat{\alpha}_t = \mathcal{N}(\boldsymbol{x}^{\mathrm{b}}(t)|0, \boldsymbol{C}_{\mathrm{f}}(t)) \tag{E.1.11}$$

and $\boldsymbol{C}_{\mathrm{f}}(t) = \langle \boldsymbol{x}^{\mathrm{b}}(t)\boldsymbol{x}^{\mathrm{b}}(t)^T \rangle$ is the equal time forward (or "filtered") posterior covariance.

By substituting (E.1.5), (E.1.6) and (E.1.11) into (E.1.10) and identifying the quadratic terms in $\boldsymbol{x}^{\mathrm{b}}(t)$ in the exponents one obtains the recursive Kalman filter expression for $\boldsymbol{C}_{\mathrm{f}}^{-1}(t)$

$$\boldsymbol{C}_{\mathrm{f}}^{-1}(t) \quad = \quad \left[(\mathbb{1} + \Delta\boldsymbol{K}^{\mathrm{bb}})\,\boldsymbol{C}_{\mathrm{f}}(t-\Delta)(\mathbb{1} + \Delta\boldsymbol{K}^{\mathrm{bb}})^T + \Delta\boldsymbol{\Sigma}^{\mathrm{bb}}\right]^{-1} + \Delta\boldsymbol{W} \tag{E.1.12}$$

where $\boldsymbol{W} = \boldsymbol{K}^{\mathrm{sb}\,T}(\boldsymbol{\Sigma}^{\mathrm{ss}})^{-1}\boldsymbol{K}^{\mathrm{sb}}$ is the *feedback* matrix. Equation (E.1.12) is a discrete time Riccati (i.e. second order matrix) recursion. We are interested in the continuous time limit $\Delta \to 0$, where it becomes

$$\frac{d}{dt}\boldsymbol{C}_{\mathrm{f}}^{-1}(t) = \boldsymbol{C}_{\mathrm{f}}^{-1}(t)\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{f}}^{-1}(t) + \boldsymbol{C}_{\mathrm{f}}^{-1}(t)\boldsymbol{K}^{\mathrm{bb}} + \boldsymbol{K}^{\mathrm{bb}\,T}\boldsymbol{C}_{\mathrm{f}}^{-1}(t) + \boldsymbol{W} \tag{E.1.13}$$

From (E.1.13) an equation for $\boldsymbol{C}_{\mathrm{f}}(t)$ can be immediately derived

$$\frac{d}{dt}\boldsymbol{C}_{\mathrm{f}}(t) = \boldsymbol{\Sigma}^{\mathrm{bb}} + \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{f}}(t) + \boldsymbol{C}_{\mathrm{f}}(t)\boldsymbol{K}^{\mathrm{bb}\,\mathrm{T}} - \boldsymbol{C}_{\mathrm{f}}(t)\boldsymbol{W}\boldsymbol{C}_{\mathrm{f}}(t) \tag{E.1.14}$$

The stationary value $\boldsymbol{C}_{\mathrm{f}}$ is described by the condition

$$\boldsymbol{\Sigma}^{\mathrm{bb}} + \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{f}} + \boldsymbol{C}_{\mathrm{f}}\boldsymbol{K}^{\mathrm{bb}\,\mathrm{T}} - \boldsymbol{C}_{\mathrm{f}}\boldsymbol{W}\boldsymbol{C}_{\mathrm{f}} = 0 \tag{E.1.15}$$

An exhaustive discussion regarding the existence and structure of solutions as well as methods to constructively find them for Riccati algebraic equations can be found in [113]. Conditions of convergence and of uniqueness of solutions for the Riccati recursion (E.1.14) have also been systematically assessed in connection to system properties (see [12, 96]). We shall briefly focus

on the condition guaranteeing the convergence of the Riccati recursion (E.1.14) to a stable time invariant solution given by (E.1.15), i.e. the *stabilizability* of $[(1+\Delta \boldsymbol{K}^{\mathrm{bb}}), \Delta \boldsymbol{K}^{\mathrm{sb}}]$. The two matrices $[(1+\Delta \boldsymbol{K}^{\mathrm{bb}}), \Delta \boldsymbol{K}^{\mathrm{sb}}]$ are said to be *stabilizable* if there exist a matrix $\boldsymbol{N}$ such that the eigenvalues of $(1+\Delta \boldsymbol{K}^{\mathrm{bb}}) - \boldsymbol{N}\boldsymbol{K}^{\mathrm{sb}}$ lie inside the unit disk. The matrix $\boldsymbol{N}$ is known as *Kalman gain* or *"observer" gain* matrix and it intervenes in the forward computation of the conditional mean as correction coefficient which embeds the contribution of observations. In this context it turns out to play a stabilizing role for the conditional forward dynamics. In the small Δ limit, $\boldsymbol{N} \sim \Delta \boldsymbol{C}_{\mathrm{f}} \boldsymbol{K}^{\mathrm{sb}\,\mathrm{T}} (\boldsymbol{\Sigma}^{\mathrm{ss}})^{-1}$ so that $\boldsymbol{N}\boldsymbol{K}^{\mathrm{sb}} \sim \Delta \boldsymbol{C}_{\mathrm{f}} \boldsymbol{W}$. The requirement of $(1+\Delta \boldsymbol{K}^{\mathrm{bb}}) - \Delta \boldsymbol{C}_{\mathrm{f}} \boldsymbol{W}$ eigenvalues to be inside the unit disk, i.e.

$$|1 + \Delta(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{C}_{\mathrm{f}} \boldsymbol{W})| < 1 \tag{E.1.16}$$

implies that $(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{C}_{\mathrm{f}} \boldsymbol{W})$ is negative definite. We thus can verify formally what would be expected by construction: first, in absence of observations, i.e. $\boldsymbol{W} \equiv 0$, the bulk internal stability condition can be retrieved (as it coincides with $\boldsymbol{K}^{\mathrm{bb}}$ being a negative matrix). Next, if the latter holds true, the condition of $(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{C}_{\mathrm{f}} \boldsymbol{W})$ being negative definite is trivially satisfied, as both $\boldsymbol{C}_{\mathrm{f}}$ and $\boldsymbol{W}$ are positive matrices: as a consequence, the internal bulk stability always implies the conditional one but the reverse is not always true. As stressed in section 4.3.3, a finite conditional variance could be in principle compatible with a divergent hidden dynamics: we nevertheless neglect these cases mirroring unphysical situations one is not typically interested in. For example, in the case of just self-interactions, i.e. $\boldsymbol{K}^{\mathrm{bb}} = -\lambda \mathbb{1}$, (E.1.15) becomes

$$\sigma_{\mathrm{b}}^2 - 2\lambda c^{\mathrm{f}} - (c^{\mathrm{f}})^2 w = 0 \tag{E.1.17}$$

whose positive definite solution is

$$c^{\mathrm{f}} = \frac{-\lambda + \sqrt{\lambda^2 + \sigma_{\mathrm{b}}^2 w}}{w} \tag{E.1.18}$$

The stability condition $-\lambda - c^{\mathrm{f}} w < 0$ reads

$$-\lambda - \Big(\frac{-\lambda + \sqrt{\lambda^2 + \sigma_{\mathrm{b}}^2 w}}{w}\Big) w = -\sqrt{\lambda^2 + \sigma^2 \hat{w}} < 0 \tag{E.1.19}$$

which is trivially satisfied except for $\lambda/\sigma \to 0$ when $0 \leq \alpha \leq 1$ (as only for these α values $\hat{w}$ can be zero): in this way we simply recover the stability condition stated in section 4.3.2.

The backward propagation incorporates in the algorithm the observations from all later time steps

$$\hat{\beta}_t \propto \int d\boldsymbol{x}^{\mathrm{b}}(t+\Delta) \hat{\beta}_{t+\Delta} P(\boldsymbol{x}^{\mathrm{s}}(t+\Delta)|\boldsymbol{x}^{\mathrm{b}}(t+\Delta)) P(\boldsymbol{x}^{\mathrm{b}}(t+\Delta)|\boldsymbol{x}^{\mathrm{b}}(t)) \tag{E.1.20}$$

and we set

$$\hat{\beta}_t \propto \mathcal{N}(\boldsymbol{x}^{\mathrm{b}}(t)|0, \boldsymbol{C}_{\mathrm{b}}(t)) \tag{E.1.21}$$

with $\boldsymbol{C}_{\rm b}(t) = \langle \boldsymbol{x}^{\rm b}(t)\boldsymbol{x}^{\rm b}(t)^T \rangle$ defined as the equal time posterior variance in the backward propagation. Inserting (E.1.21) into (E.1.20) one finds the backward recursion for $\boldsymbol{C}_{\rm b}^{-1}(t)$

$$\boldsymbol{C}_{\rm b}^{-1}(t) = (\mathbb{1} + \Delta \boldsymbol{K}^{\rm bb})^T (\Delta \boldsymbol{\Sigma}^{\rm bb})^{-1} [\mathbb{1} - (\mathbb{1} + \Delta \boldsymbol{\Sigma}^{\rm bb} \boldsymbol{C}_{\rm b}^{-1}(t+\Delta) + \Delta^2 \boldsymbol{\Sigma}^{\rm bb} \boldsymbol{W})^{-1}](\mathbb{1} + \Delta \boldsymbol{K}^{\rm bb})$$

Taking $\Delta \to 0$, which requires keeping all terms up to $O(\Delta)$ on the r.h.s., gives the continuous time limit

$$\frac{d}{dt}\boldsymbol{C}_{\rm b}^{-1}(t) = -\boldsymbol{K}^{\rm bb\,T}\boldsymbol{C}_{\rm b}^{-1}(t) - \boldsymbol{C}_{\rm b}^{-1}(t)\boldsymbol{K}^{\rm bb} - \boldsymbol{W} + \boldsymbol{C}_{\rm b}^{-1}(t)\boldsymbol{\Sigma}^{\rm bb}\boldsymbol{C}_{\rm b}^{-1}(t) \tag{E.1.22}$$

The changes of sign compared to (E.1.13) come from the backward direction.

Finally the posterior γ_t also has a Gaussian form,

$$\gamma_t = \mathcal{N}(\boldsymbol{x}^{\rm b}(t)|0, \boldsymbol{C}^{\rm bb|s}(t)) \tag{E.1.23}$$

We drop the superscripts on $\boldsymbol{C}^{\rm bb|s}(t)$ as in the main text and write this overall ("smoothed") covariance as $\boldsymbol{C}(t)$. From (E.1.9) one has $\boldsymbol{C}^{-1}(t) = \boldsymbol{C}_{\rm f}^{-1}(t) + \boldsymbol{C}_{\rm b}^{-1}(t)$, so from the sum of (E.1.13) and (E.1.22)

$$\frac{d}{dt}\boldsymbol{C}^{-1}(t) = \boldsymbol{C}^{-1}(t)\boldsymbol{\Sigma}^{\rm bb}\boldsymbol{C}^{-1}(t) + \boldsymbol{C}^{-1}(t)\boldsymbol{K}^{\rm bb|s} + \boldsymbol{K}^{\rm bb|s\,T}\boldsymbol{C}^{-1}(t) \tag{E.1.24}$$

where we have set

$$\boldsymbol{K}^{\rm bb|s} = \boldsymbol{K}^{\rm bb} - \boldsymbol{\Sigma}^{\rm bb}\boldsymbol{C}_{\rm b}^{-1} \tag{E.1.25}$$

and we have taken $\boldsymbol{C}_{\rm b}^{-1}$ as the stationary limit of $\boldsymbol{C}_{\rm b}^{-1}(t)$.

To interpret $\boldsymbol{K}^{\rm bb|s}$ one can look at $P(\boldsymbol{x}^{\rm b}(t+\Delta), \boldsymbol{x}^{\rm b}(t)|\boldsymbol{X}^{\rm s})$, given by the integrand of (E.1.20). Conditioning on $\boldsymbol{x}^{\rm b}(t)$ and using (E.1.5), (E.1.6) and (E.1.21) one finds easily

$$P(\boldsymbol{x}^{\rm b}(t+\Delta), \boldsymbol{x}^{\rm b}(t)|\boldsymbol{X}^{\rm s}) \sim \tag{E.1.26}$$
$$e^{\boldsymbol{x}^{\rm b\,T}(t+\Delta)\boldsymbol{C}_{\rm b}^{-1}(t+\Delta)\boldsymbol{x}^{\rm b}(t+\Delta)} e^{(\boldsymbol{x}^{\rm b}(t+\Delta) - \boldsymbol{x}^{\rm b}(t) - \Delta\boldsymbol{K}^{\rm bb}\boldsymbol{x}^{\rm b}(t))^T (\Delta\boldsymbol{\Sigma}^{\rm bb})^{-1} (\boldsymbol{x}^{\rm b}(t+\Delta) - \boldsymbol{x}^{\rm b}(t) - \Delta\boldsymbol{K}^{\rm bb}\boldsymbol{x}^{\rm b}(t))} e^{(\Delta\boldsymbol{K}^{\rm sb}\boldsymbol{x}^{\rm b}(t))^T (\Delta\boldsymbol{\Sigma}^{\rm ss})^{-1} (\Delta\boldsymbol{K}^{\rm sb}\boldsymbol{x}^{\rm b}(t))}$$

thus the mean of $\boldsymbol{x}^{\rm b}(t+\Delta)$ conditioned on $\boldsymbol{x}^{\rm b}(t)$ is

$$(\mathbb{1} + \Delta\boldsymbol{K}^{\rm bb|s}(t) + O(\Delta^2))\boldsymbol{x}^{\rm b}(t) \tag{E.1.27}$$

Hence $\boldsymbol{K}^{\rm bb|s}(t)$ has the meaning of a posterior drift, i.e. it determines the time evolution for the posterior dynamics.

Focussing on the stationary state now, we can drop all dependences on t. From (E.1.24), the posterior covariance $\boldsymbol{C}$ then satisfies the Lyapunov equation (4.2.5)

$$\boldsymbol{K}^{\rm bb|s}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\rm bb|s\,T} + \boldsymbol{\Sigma}^{\rm bb} = 0 \tag{E.1.28}$$

with the stationary posterior drift $\boldsymbol{K}^{\mathrm{bb|s}}$ given by

$$\boldsymbol{K}^{\mathrm{bb|s}} = \boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{b}}^{-1} \tag{E.1.29}$$

and the stationary backward covariance satisfying, from (E.1.22)

$$\boldsymbol{C}_{\mathrm{b}}^{-1}\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{b}}^{-1} - \boldsymbol{K}^{\mathrm{bb}\,T}\boldsymbol{C}_{\mathrm{b}}^{-1} - \boldsymbol{C}_{\mathrm{b}}^{-1}\boldsymbol{K}^{\mathrm{bb}} = \boldsymbol{W} \tag{E.1.30}$$

Apart from the relabelling of $\boldsymbol{C}_{\mathrm{b}}^{-1}$ as $\boldsymbol{A}$, we have therefore derived (4.2.5), (4.2.7) and (4.2.8) in the main text.

To find the evolution of the two-time posterior variance $\boldsymbol{C}(t, t')$, we first look at the case $\boldsymbol{C}(t' + \Delta, t')$ of adjacent time steps. Here (E.1.27) gives directly

$$\boldsymbol{C}(t' + \Delta, t') = (\mathbb{1} + \Delta\boldsymbol{K}^{\mathrm{bb|s}}(t') + O(\Delta^2))\boldsymbol{C}(t', t'). \tag{E.1.31}$$

This easily generalizes to the correlations τ steps apart as

$$\boldsymbol{C}(t' + \tau\Delta, t') = (\mathbb{1} + \Delta\boldsymbol{K}^{\mathrm{bb|s}} + O(\Delta^2))^{\tau}\boldsymbol{C} \tag{E.1.32}$$

where we have directly written the stationary version. Setting $t = t' + \tau\Delta$ and taking $\Delta \to 0$ then gives equation (4.2.9) in the main text, i.e.

$$\boldsymbol{C}(t - t') = \lim_{\Delta\to 0}\left[\left(\mathbb{1} + \Delta\boldsymbol{K}^{\mathrm{bb|s}}\frac{(t - t')}{(t - t')}\right)\right]^{\frac{t-t'}{\Delta}}\boldsymbol{C} = e^{\boldsymbol{K}^{\mathrm{bb|s}}(t-t')}\boldsymbol{C} \tag{E.1.33}$$

where $(1 + x/n)^n \to e^x$ for $n \to \infty$ has been used. We see that the 2-time posterior covariance decays exponentially and the rate of such a decay is determined by the posterior drift $\boldsymbol{K}^{\mathrm{bb|s}}$, which should be negative definite to ensure the stability of the posterior dynamics. If the hidden trajectories do not diverge, i.e. $\boldsymbol{K}^{\mathrm{bb}}$ is a negative matrix, the condition of $\boldsymbol{K}^{\mathrm{bb|s}} = \boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{C}_{\mathrm{b}}^{-1}$ being negative definite is trivially satisfied, as both $\boldsymbol{\Sigma}^{\mathrm{bb}}$ and $\boldsymbol{C}_{\mathrm{b}}^{-1}$ are symmetric by definition and positive semi-definite. As their product enters the effective drift with a minus sign, we see that the presence of observations drives the hidden dynamics back towards its mean (zero) more quickly.

In absence of observations, i.e. $\boldsymbol{W} = 0$, we note $\boldsymbol{C}_{\mathrm{b}}^{-1}(t) \equiv 0\ \forall t$, so that $\boldsymbol{C}(t) \equiv \boldsymbol{C}_{\mathrm{f}}(t)\ \forall t$ and $\boldsymbol{K}^{\mathrm{bb|s}} \equiv \boldsymbol{K}^{\mathrm{bb}}$ as expected: (E.1.28) reduces to a Lyapunov equation for $\boldsymbol{C}$ containing the simple drift $\boldsymbol{K}^{\mathrm{bb}}$. As a consequence, the internal bulk stability ($\boldsymbol{K}^{\mathrm{bb}}$ negative definite) always implies the conditional one but the reverse is not always true, as already pointed out for the filtering step.

Finally, we see from (4.2.7) and (E.1.25) that $\boldsymbol{C}_{\mathrm{b}}^{-1} = \boldsymbol{A}$, $\boldsymbol{A}$ being the Lagrange multiplier we introduce in the variational derivation of appendix F.

Appendix F

Variational method

F.1 Variational method

As is often the case, the fixed point of a recursion (such as the Forward-Backward algorithm) can also be retrieved variationally, i.e. as the solution of a constrained optimization problem. We show this connection in this appendix.

Let us start from $P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}})$, the joint probability of subnetwork and bulk trajectories obeying (4.2.1a) and (4.2.1b), and denote $Q(\boldsymbol{X}^{\mathrm{b}})$ a variational approximation to the posterior $P(\boldsymbol{X}^{\mathrm{b}}|\boldsymbol{X}^{\mathrm{s}})$ of the effective dynamics (4.2.6). As before if we are interested only in the posterior second order statistics, we can remove the means by assuming $\boldsymbol{x}^{\mathrm{s}}(t) = 0\ \forall t$ and can then drop the δ in (4.2.6) (by general theory of Gaussian conditional processes, the conditional mean $\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t)$ is linearly related to observations, see in the following appendix I, section I.1). One aim is to determine the effective drift $\boldsymbol{K}^{\mathrm{bb}|\mathrm{s}}$ by variational methods. Note that parameterizing Q in terms of $\boldsymbol{K}^{\mathrm{bb}|\mathrm{s}}$ gives us enough flexibility to retrieve the *exact* posterior because of the Gaussian nature of our problem. We can write the joint trajectory probability and the variational posterior, directly in continuous time form, as

$$P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}}) \propto \exp\left[-\frac{1}{2}\int_0^T dt(\boldsymbol{\xi}^{\mathrm{b}\,T}(t)\boldsymbol{\Sigma}^{\mathrm{bb}\,-1}\boldsymbol{\xi}^{\mathrm{b}}(t) + \boldsymbol{\xi}^{\mathrm{s}\,T}(t)\boldsymbol{\Sigma}^{\mathrm{ss}\,-1}\boldsymbol{\xi}^{\mathrm{s}}(t))\right] \tag{F.1.1}$$

$$Q(\boldsymbol{X}^{\mathrm{b}}) \propto \exp\left[-\frac{1}{2}\int_0^T dt\,\boldsymbol{\xi}^{\mathrm{b}\,T}(t)\boldsymbol{\Sigma}^{\mathrm{bb}\,-1}\boldsymbol{\xi}^{\mathrm{b}}(t)\right] \tag{F.1.2}$$

where the noises $\boldsymbol{\xi}^{\mathrm{b}}$ and $\boldsymbol{\xi}^{\mathrm{s}}$ should be expressed as a function of $\boldsymbol{x}^{\mathrm{b}}$ and $\boldsymbol{x}^{\mathrm{s}}$ using respectively equations (4.2.1a) and (4.2.1b) for $P(\boldsymbol{X}^{\mathrm{b}}, \boldsymbol{X}^{\mathrm{s}})$ and (4.2.6) for $Q(\boldsymbol{X}^{\mathrm{b}})$.

We find Q in the standard variational way by finding the stationary point of the Kullback-Leibler divergence [114] between P and Q

$$\mathrm{KL}(P\|Q) = -\left\langle \log \frac{Q}{P} \right\rangle_Q = F \tag{F.1.3}$$

which is analogous to a thermodynamic free energy F. Inserting (F.1.1) and (F.1.2) and simplifying gives

$$F = \frac{1}{2}\int_0^T dt\, \left\langle \boldsymbol{x}^{\mathrm{b}\,T}(t)(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})^T \boldsymbol{\Sigma}^{\mathrm{bb}\,-1}(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})\boldsymbol{x}^{\mathrm{b}}(t)\right\rangle_Q + \frac{1}{2}\int_0^T dt\, \left\langle \boldsymbol{x}^{\mathrm{b}\,T}(t)\boldsymbol{W}\boldsymbol{x}^{\mathrm{b}}(t)\right\rangle_Q \tag{F.1.4}$$

with $\boldsymbol{W} \doteq (\boldsymbol{K}^{\mathrm{sb}})^T \boldsymbol{\Sigma}^{\mathrm{ss}\,-1}\boldsymbol{K}^{\mathrm{sb}}$ the feedback matrix as before. Here we have performed an integration by parts and assumed that $\boldsymbol{x}^{\mathrm{b}}$ vanishes at the boundaries T of the time domain.

In the stationary limit, we can drop the time integrals, drop the resulting factor T and use the definition $\boldsymbol{C} = \langle \boldsymbol{x}^{\mathrm{b}}\boldsymbol{x}^{\mathrm{b}\,T}\rangle_Q$ to write

$$F = \frac{1}{2}\mathrm{Tr}\left[(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})^T \boldsymbol{\Sigma}^{\mathrm{bb}\,-1}(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})\boldsymbol{C}\right] + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{W}\boldsymbol{C}\right] \tag{F.1.5}$$

We now want to optimize over $\boldsymbol{K}^{\mathrm{bb|s}}$, bearing in mind that the stationary posterior variance $\boldsymbol{C}$ is linked to the effective drift by the Lyapunov equation

$$\boldsymbol{K}^{\mathrm{bb|s}}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\mathrm{bb|s}\,T} + \boldsymbol{\Sigma}^{\mathrm{bb}} = 0 \tag{F.1.6}$$

(see (4.2.5) in the main text). Introducing a Lagrange multiplier matrix $\boldsymbol{A}/2$ to implement this constraint, we optimize

$$\mathcal{L}[\boldsymbol{C}, \boldsymbol{K}^{\mathrm{bb|s}}, \boldsymbol{A}] = F + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{A}^T(\boldsymbol{K}^{\mathrm{bb|s}}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\mathrm{bb|s}\,T} + \boldsymbol{\Sigma}^{\mathrm{bb}})\right] = \tag{F.1.7}$$
$$\frac{1}{2}\mathrm{Tr}\left[(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})^T \boldsymbol{\Sigma}^{\mathrm{bb}\,-1}(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{K}^{\mathrm{bb|s}})\boldsymbol{C}\right] + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{W}\boldsymbol{C}\right] + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{A}^T(\boldsymbol{K}^{\mathrm{bb|s}}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{K}^{\mathrm{bb|s}\,T} + \boldsymbol{\Sigma}^{\mathrm{bb}})\right]$$

Optimization w.r.t. $\boldsymbol{K}^{\mathrm{bb|s}}$ gives

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{K}^{\mathrm{bb|s}}} = \boldsymbol{\Sigma}^{\mathrm{bb}\,-1}(\boldsymbol{K}^{\mathrm{bb|s}} - \boldsymbol{K}^{\mathrm{bb}})\boldsymbol{C} + \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{C} = 0 \tag{F.1.8}$$

from which one has the expression (4.2.7) for the posterior drift matrix; in fact

$$\boldsymbol{K}^{\mathrm{bb|s}} = \boldsymbol{K}^{\mathrm{bb}} - \frac{\boldsymbol{\Sigma}^{\mathrm{bb}}}{2}(\boldsymbol{A} + \boldsymbol{A}^T) = \boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A}_{\mathrm{s}} \tag{F.1.9}$$

where we have denoted the symmetric part of $\boldsymbol{A}$ by $\boldsymbol{A}_{\mathrm{s}} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T)$. We will then write $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{s}} + \boldsymbol{A}_{\mathrm{a}}$ with $\boldsymbol{A}_{\mathrm{a}} = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)$ the antisymmetric part. The second optimization condition reads

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{C}} = \frac{1}{2}(\boldsymbol{K}^{\mathrm{bb|s}} - \boldsymbol{K}^{\mathrm{bb}})^T \boldsymbol{\Sigma}^{\mathrm{bb}\,-1}(\boldsymbol{K}^{\mathrm{bb|s}} - \boldsymbol{K}^{\mathrm{bb}}) + \frac{1}{2}\boldsymbol{W} + \frac{1}{2}(\boldsymbol{A}\boldsymbol{K}^{\mathrm{bb|s}} + \boldsymbol{K}^{\mathrm{bb|s}\,T}\boldsymbol{A}) = 0 \tag{F.1.10}$$

By substitution of (F.1.9) into (F.1.10) one obtains

$$\boldsymbol{A}_{\mathrm{s}}\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A}_{\mathrm{s}} - \boldsymbol{K}^{\mathrm{bb}\,T}\boldsymbol{A}_{\mathrm{s}} - \boldsymbol{A}_{\mathrm{s}}\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{A}_{\mathrm{a}}(\boldsymbol{K}^{\mathrm{bb}} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A}_{\mathrm{s}}) - (\boldsymbol{K}^{\mathrm{bb}\,T} - \boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A}_{\mathrm{s}})\boldsymbol{A}_{\mathrm{a}} - \boldsymbol{W} = 0$$

The symmetric part of this determines $\boldsymbol{A}_{\mathrm{s}}$, which is all we need for (F.1.9), as

$$\boldsymbol{A}_{\mathrm{s}}\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{A}_{\mathrm{s}} - \boldsymbol{K}^{\mathrm{bb}\,T}\boldsymbol{A}_{\mathrm{s}} - \boldsymbol{A}_{\mathrm{s}}\boldsymbol{K}^{\mathrm{bb}} = \boldsymbol{W} \tag{F.1.11}$$

This is equation (4.2.8) in the main text – we dropped the subscript "s" there – and shows that the Lagrange multiplier $\boldsymbol{A}$ is identical to the (stationary) inverse backward covariance matrix, $\boldsymbol{C}_{\mathrm{b}}^{-1}$.

The above derivation signifies that the Lagrange multiplier $\boldsymbol{A}$ plays the role of the inverse stationary variance computed by backward propagation: in light of either interpretations, it is expected to vanish in absence of constraints by observations.

Finally $\partial\mathcal{L}/\partial\boldsymbol{A} = 0$ simply ensures that $\boldsymbol{C}$ satisfies (F.1.6) with $\boldsymbol{K}^{\mathrm{bb|s}}$ given by (F.1.9).

CHAPTER 5

Gaussian Variational Approximation for biochemical networks

The related paper will be submitted to Physical Biology [115].

> Che nel gioco dei dadi alcuni punti sieno più vantaggiosi di altri, vi ha la sua ragione assai manifesta, la quale è, il poter quelli più facilmente e più frequentemente scoprirsi, che questi, il che dipende dal potersi formare con più sorte di numeri.[1]
>
> *Galileo Galilei, "Sopra le scoperte dei dadi"*

5.1 Introduction

Protein-protein interaction networks, such as the ones building up signalling pathways, can contain thousands of reacting species and the underlying mechanisms are still largely unclear [116, 117]. With numerous, nonlinear equations to solve we need to resort to approximate descriptions and model reduction techniques, as we discussed in chapter 1.

Sampling-based approaches to the Chemical Master Equation [11, 12] are powerful tools to deal with the complexity of biochemical reactions; still they often need in parallel some clever sim-

[1] The fact that in a dice-game certain numbers are more advantageous than others has a very obvious reason, i.e. that some are more easily and more frequently made than others, which depends on their being able to be made up with more variety of numbers.

plification to be implemented in a less computationally expensive way. Analytic approximations can be put forward as an extremely valuable tool in this respect, while giving further insight into the underlying processes. Our starting point for this aim will be the Chemical Langevin Equation [10], a diffusion-type approximation of the Master Equation.

Variational approaches can be used to obtain efficient approximations of probability distributions over the temporal trajectories (paths) for stochastic dynamics; the key element is a parametrized ansatz motivated by some available knowledge of system. Eyink [118] has built a general scheme to produce approximations by applying a variational principle on a non-equilibrium effective action; see [119, 120] for applications relying on Poisson distributions as ansatz.

This same problem of approximating intractable distributions (either marginals or posterior) can be systematically tackled via machine learning techniques [16]. Among them, the minimization of the Kullback-Leibler divergence [114] is a tool to obtain the most faithful approximation within a chosen family of approximating distributions; examples are the mean field-type (factorized) ansatz [80, 121] or the choice of Gaussian measures [37, 110], the simplest approximation beyond mean field. In this chapter, we resort to a Gaussian variational approximation of the paths distribution for stochastic equations describing networks of unary and binary biochemical reactions. The Gaussian assumption clearly yields a second order moment closure scheme, first outlined by Whittle [122], for which several precedents do exist in modelling biochemical reactions [123, 124] and whose performance has been studied in comparison to other approximations [125–127].

For the sake of realistic descriptions, it is crucial also to capture the extrinsic factors modulating the dynamics of particular reaction networks, in order to mirror the significant environmental fluctuations detected in *in vivo* conditions. Deriving models for subsystems embedded into a larger environment in a *systematic* and *principled* way can therefore serve more efficiently the modelling purpose and also the inference one. In fact, assuming that some knowledge of the subsystem can be established, one could think of fixing its variables in the model and work out the relations characterizing the inference of environmental features. To attain this, it is necessary to reduce the problem to equations for fewer species with the minimal loss of mesoscopic information, i.e. in such a way that the solution still retains dynamical properties of the whole ensemble. As we have discussed in chapter 1, this problem has motivated a noticeable effort in developing model reduction strategies for systems biology [1, 2, 5, 6], model reduction being an effective tool for simplification and better understanding more widely in complex systems. An established approach towards this goal is given by projection methods, which were first introduced in irreversible statistical mechanics [128] and more recently have obtained attention in the field of model reduction

[77, 129–131]. The setting we consider in this chapter is the following: we focus on a subset of species, what we refer to as "subnetwork", which is embedded in a larger network, the "bulk", which we assume to be under-resolved with respect to the subnetwork. The subnetwork is selected either because experimentally entirely accessible or because better characterized from the theoretical point of view or because more relevant for the biological features of interest. To complement the analysis of chapters 3 and 4, where the focus was on the inverse problem of inferring bulk dynamics, in this chapter we tackle especially a *direct* problem. Our purpose is to calculate the evolution of subnetwork variables without knowing every detail of the full network, but starting from just some prior information, in the form of the statistics of bulk initial conditions, the subnetwork-bulk and bulk-bulk interactions.

Projection methods show explicitly that the structure of such a reduced dynamics includes a memory term and an additional random force, as we will explain in section 5.2.1; the main drawback resides in finding closed form expressions for them. Including non-Markovian temporal effects yields the appropriate coarse-grained description and a higher accuracy of prediction in different contexts, e.g. harmonic systems [132] and colloidal particles [133]. Interestingly, Rubin et al. [77, 134] worked out the projection formalism for the unary and binary reactions with mass action kinetics we take into consideration here and also extended them to Michaelis-Menten kinetics [135]. We regard these previous works as essential terms of comparison and consistency tests for our results. The novel contribution of this chapter is a model reduction method based on a Gaussian approximation which is capable of accurately capturing nonlinearities of quadratic type. As for [77], our method is applicable to *arbitrarily chosen* subnetworks, while other path-based coarse graining procedures are tailored to merge or eliminate only fast reactions [136]. Instead of dealing with projection operators, we derive the stochastic equations reduced to the subnetwork by *marginalizing* bulk degrees of freedom (d.o.f.) from a path integral representation of the full dynamics.

The *marginalization* of a joint process w.r.t. extrinsic components is the key concept to uncouple a generic subnetwork from its environment in such a way as to still incorporate the dynamical effects of the latter. This has been highlighted in other recent works analyzing fixed but random environmental conditions [137] and fluctuating ones [138]. Importantly, here we will consider a case where the environment (bulk), equivalently to the subnetwork, is modelled as a network, i.e. with its own structure and dynamics. This fact ensures more flexibility in choosing and interchanging subnetworks and the embedding bulk and can be thus useful to adhere to different datasets.

This chapter is organized as follows. We first present the model and the structure of its subnetwork-

reduced version as expected from projection methods in section 5.2. In section 5.3 we revisit a variational derivation of a Gaussian approximation over paths [110], leading to an effective linearization of the dynamics (we provide the details of this calculation in the appendix G). We derive the additional terms in the subnetwork dynamics that arise via the bulk, i.e. a memory and a coloured noise, first for this *linearized* dynamics (section 5.4.1); next, writing the reduced subnetwork dynamics in terms of an "effective" (i.e. where the bulk is integrated out) action allows us to perform a perturbative expansion, in such a way that the *nonlinear* memory and noise can be read off by inspection of this effective action (section 5.5.1 and appendix I). In section 5.5.3 and appendix J, we show that the resulting approximation is equivalent to the projected equations of [77] and we argue its advantages w.r.t. the latter in terms of computational expediency by discussing its numerical implementation (section 5.5.2 and appendix K). In section 5.6.1 we consider a toy model, instructive in its simplicity, for illustrating the method and for testing its accuracy in contrast to usual approximation schemes. Finally, in section 5.6.2, we apply our model reduction to the protein-protein interaction network around Epidermal Growth Factor Receptor (EGFR) to show the increased accuracy of prediction it yields.

5.2 Set up

We begin with a set of Chemical Langevin Equation (CLE) [10] for concentrations of proteins (labelled by indexes $i \in \{1...N\}$)

$$\frac{\partial x_i(t)}{\partial t} = \Phi_i(\boldsymbol{x}(t)) + \xi_i(t) \tag{5.2.1}$$

where $\xi_i(t)$ is an *intrinsic* noise term (manifestation of copy number fluctuations) and

$$\begin{aligned}\Phi_i(\boldsymbol{x}(t)) = &\sum_{j,l,j\neq l}\left(k^-_{l,ij}x_i(t) - k^+_{ij,l}x_i(t)x_j(t)\right) + \frac{1}{2}\sum_{j,l,j\neq l}\left(k^+_{jl,i}x_j(t)x_l(t) - k^-_{i,jl}x_i(t)\right) + \\ &\sum_j\left(\lambda_{ji}x_j(t) - \lambda_{ij}x_i(t)\right) + \sum_l\left(2k^-_{l,ii}x_l(t) - k^+_{ii,l}x_i(t)x_i(t)\right) + \sum_j\left(\frac{1}{2}k^+_{jj,i}x_j(t)x_j(t) - k^-_{i,jj}x_i(t)\right)\end{aligned} \tag{5.2.2}$$

These dynamical equations capture the formation of complex l from proteins i and j (with rate $k^+_{ij,l}$), the dissociation of complex l into proteins i and j ($k^-_{l,ij}$), and the change of protein i into protein j (λ_{ij}), as schematically shown

- Complex formation $\quad i + j \xrightarrow{k^+_{ij,l}} l$
- Complex dissociation $\quad l \xrightarrow{k^-_{l,ij}} i + j$
- Conformation change $\quad i \xrightarrow{\lambda_{ij}} j$

We are considering the formation both of heterodimers and homodimers. A CLE-based description is equivalent to the Chemical Fokker Planck equation, which is obtained as truncation of Kramers-Moyal expansion of the Master Equation after the first order in $1/V$ [27]. By construction of the CLE [10], the noise is Gaussian distributed with zero mean and covariance matrix

$$\langle \xi_i(t)\xi_j(t')\rangle = \Sigma_{ij}(\boldsymbol{x}, t, t') = \Sigma_{ij}(\boldsymbol{x})\delta(t-t') \tag{5.2.3}$$

thus uncorrelated in time (white noise) but multiplicative (with $\boldsymbol{x}$-dependent correlations). (We appeal to the Itô convention [71] for this stochastic term). One can show [64] that, in the CLE approximation, $\boldsymbol{\Sigma}(\boldsymbol{x}(t))$ can be written as

$$\boldsymbol{\Sigma}(\boldsymbol{x}(t)) = \frac{\epsilon}{2}\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{f}(\boldsymbol{x}(t)))\,\boldsymbol{S}^T \tag{5.2.4}$$

where $\boldsymbol{S}$ denotes the stoichiometric matrix and $\boldsymbol{f}$ the flux vector (containing reaction rate constants). $\boldsymbol{S}$ and $\boldsymbol{f}$ encode the deterministic part of Langevin dynamics (5.2.1), which can be then be rewritten $\partial_t \boldsymbol{x}(t) = \boldsymbol{S}\boldsymbol{f}(\boldsymbol{x}(t)) + \boldsymbol{\xi}(t)$. Finally ϵ is a parameter setting the amplitude of fluctuations and is given by $\epsilon = 1/V$, V being the reaction volume (typically the cell volume): the noise contribution diminishes as the volume V increases. This relation follows the fact that the molecules' copy number N is Poisson-distributed, i.e. $N \approx \mathrm{Poisson}(Vx)$ with x indicating the concentration. Fluctuations on the copy number are then

$$\frac{\langle(\Delta N)^2\rangle}{\langle N\rangle^2} = \frac{1}{Vx} \approx \frac{1}{V} \tag{5.2.5}$$

as $\langle N\rangle = Vx$ and $\langle(\Delta N)^2\rangle = \langle N\rangle = Vx$ (property of Poisson distributions).

Following the approach in [139], we define a generating functional as in the Martin–Siggia–Rose–Janssen–De Dominicis (MSRJD) functional integral formalism [19–21], of the time-dependent statistics of protein concentrations (see also [75] for a systematic explanation of this procedure). First of all, we discretize time, the fundamental time step being Δ, and we use a functional $\delta-$function to impose the validity of equation (5.2.1). The Fourier representation of the δ leads us to introduce a set of auxiliary variables $\hat{\boldsymbol{x}}$ as follows

$$\begin{aligned} Z_\xi[\boldsymbol{\psi}] &= \int \prod_{it} dx_i(t)\, e^{\mathrm{i}\,\Delta\psi_i(t)x_i(t)}\delta(x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t))) = \\ &= \int \prod_{it} \frac{dx_i(t)d\hat{x}_i(t)}{2\pi}\, e^{\mathrm{i}\Delta\psi_i(t)x_i(t)}\, e^{\mathrm{i}\hat{x}_i(t)\left[x_i(t+\Delta)-x_i(t)-\Delta\Phi_i(\boldsymbol{x}(t))-\Delta\bar{\xi}_i(t)\right]} \end{aligned} \tag{5.2.6}$$

The contribution of the noise appears as the temporal average in the time interval Δ

$$\bar{\xi}_i(t) = \frac{1}{\Delta}\int_t^{t+\Delta} \xi_i(t')dt' \tag{5.2.7}$$

so that

$$\langle \bar{\xi}_i(t)\bar{\xi}_j(t')\rangle = \Delta^{-1}\Sigma_{ij}(\boldsymbol{x})\delta(t-t') \tag{5.2.8}$$

We next average the generating functional over the noise

$$\begin{aligned} Z[\boldsymbol{\psi}] &= \langle Z_{\xi}[\boldsymbol{\psi}]\rangle_{\xi} = \\ &= \int \prod_{it} \frac{d\xi_i(t)\, dx_i(t)\, d\hat{x}_i(t)}{2\pi}\, e^{\mathrm{i}\Delta\psi_i(t)x_i(t)}\, e^{\mathrm{i}\hat{x}_i(t)\left[x_i(t+\Delta)-x_i(t)-\Delta\Phi_i(\boldsymbol{xlal}(t))-\Delta\bar{\xi}_i(t)\right]} P(\xi_i(t)) \end{aligned} \tag{5.2.9}$$

using the Hubbard-Stratonovich identity

$$\langle e^{\pm\mathrm{i}\Delta\hat{\boldsymbol{x}}\cdot\bar{\boldsymbol{\xi}}}\rangle_{\boldsymbol{\xi}} = e^{-\frac{\Delta}{2}\hat{\boldsymbol{x}}^T\boldsymbol{\Sigma}\hat{\boldsymbol{x}}} \tag{5.2.10}$$

The resulting quantity, neglecting the generating term, can be regarded as a path integral, i.e. a functional integral over an infinity of possible trajectories for $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$

$$Z[\boldsymbol{\psi}]|_{\boldsymbol{\psi}=0} = \int D\boldsymbol{x}\, D\hat{\boldsymbol{x}}\, e^{\mathcal{H}} \tag{5.2.11}$$

where we introduced $D\boldsymbol{x}\, D\hat{\boldsymbol{x}}$ as a compact notation to be read as $\prod_{it} dx_i(t)d\hat{x}_i(t)$. Every trajectory $\{x_i(t)\}$ is weighted by a factor depending on the action $\mathcal{H}$

$$\mathcal{H} = \sum_{it} \mathrm{i}\hat{x}_i(t)\left[x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t))\right] + \frac{\Delta}{2}\sum_{ijt} \mathrm{i}\hat{x}_i(t)\Sigma_{ij}(\boldsymbol{x}(t))\mathrm{i}\hat{x}_j(t) \tag{5.2.12}$$

Given this set up, our aim is to design a strategy of model reduction. In particular, we restrict our analysis to a small subset of variables, the subnetwork, and we want to derive an accurate description which accounts also for the dynamical effects of its embedding environment, the bulk.

5.2.1 Projection Methods

Projection methods are a well known tool to reformulate problems in a lower dimensional space, comprising only a subset of variables [128, 129, 140]. In this approach, the basic idea is to select relevant variables (corresponding to slow degrees of freedom) and to project the dynamical equations onto the subspace of observables spanned by them, after defining the appropriate projection operator. As a result, the contributions of the remaining less relevant variables are isolated into a *memory* kernel and a *random force* (the name is due to the fact that this term is always orthogonal to the space of relevant observables and thus uncorrelated with all their initial values). The resulting equation is known as generalized Langevin equation in irreversible statistical mechanics [141, 142] and usually describes the motion of a subsystem of particles interacting with a heat bath.

In optimal prediction methods from partial observations, one can consider the better resolved variables (i.e. the subnetwork) as the relevant observables, while the under/unresolved bulk variables are treated as the "irrelevant" ones. Briefly, if $x_i^s(t)$ is a subnetwork concentration, one can derive [140] a dynamical equation of the schematic form

$$\frac{\partial x_i^s(t)}{\partial t} = \sum_{j=1}^{N^s} x_j^s(t) \underbrace{\Omega_{ji}}_{\text{Rate matrix}} + \sum_{j=1}^{N^s} \int_0^t dt' x_j^s(t') \underbrace{M_{ji}(t-t')}_{\text{Memory function}} + \underbrace{r_i(t)}_{\text{Random force}} \tag{5.2.13}$$

where the first term Ω_{ji}, a rate matrix, is local in time and describes processes involving solely subnetwork observables. The second term, the memory, is an integral over past histories of subnetwork concentrations weighted by the memory function, which depends on the time difference. It summarizes how some dynamical features of the subnetwork, further propagated and mediated via the bulk, make the subnetwork values at time t depend on past values at times t', $t' < t$. The third term, the random force encodes unknown bulk initial conditions (thus this notion of "randomness", for its origin and justification, is distinct from the intrinsic noise). These initial conditions are assumed not to be exactly known but drawn from a probability distribution, chosen because of prior measurements or forms of prior knowledge. The statistical projection estimates their effect on the *average* evolution of selected variables, i.e. not including intrinsic noise terms. The projected solution is thus to be intended as the *conditional expectation* of the subnetwork evolution given some statistical assumption on the bulk initial values.

The projected equations of motion are formally exact, in the sense that their solution is equivalent to the average solution of the full problem, which is in this way reformulated more conveniently in terms of fewer observables. They are systematically used for complexity reduction provided that one finds the appropriate representation of the projection operator and its orthogonal space, which usually requires some approximate treatment.

Both the memory and the random force exhibit the non-trivial time dependences which fundamentally arise when switching from a description of the full network to a reduced, coarse-grained description. Including these non-Markovian effects in a reduced model yields a less significant loss of information for longer times and a more accurate prediction of time courses than any Markovian approximation, as shown in [77, 143]. In this chapter we pursue this task by deploying a variational approximation, the Gaussian Variational Approximation (GVA).

5.3 Gaussian Variational Approximation

We can exploit the path integral representation (5.2.11) to recover the dynamics of the probability distribution of protein concentrations. Such a "distribution" (more rigorously is a complex measure) of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ can be written

$$P(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{e^{\mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}}{Z} P_0(\boldsymbol{x}) \tag{5.3.1}$$

where $P_0(\boldsymbol{x})$ indicates the initial probability distribution, which is unknown.

Given that the initial dynamics (5.2.1) is nonlinear, it would be infeasible to evaluate $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$ exactly or to calculate expectations w.r.t. it. In such situations, it is necessary to resort to approximation schemes: we choose a variational inference technique. The basic idea is that an intractable distribution, such as $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$, is approximated by the closest distribution, w.r.t. a distance measure, within a family of analytically tractable ones $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$. We specifically take the approximating distribution $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ as a *Gaussian*, thus

$$P(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{e^{\mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}}{Z} P_0(\boldsymbol{x}) \approx \mathcal{N}(\boldsymbol{x}, \hat{\boldsymbol{x}} | \boldsymbol{\mu}_{\text{gen}}, \boldsymbol{C}_{\text{gen}}) = Q(\boldsymbol{x}, \hat{\boldsymbol{x}}) \tag{5.3.2}$$

The Gaussian $\mathcal{N}(\boldsymbol{x}, \hat{\boldsymbol{x}} | \boldsymbol{\mu}_{\text{gen}}, \boldsymbol{C}_{\text{gen}})$ is completely determined by two sets of parameters, the vector of mean values $\boldsymbol{\mu}_{\text{gen}}$ and the (connected) covariance matrix $\boldsymbol{C}_{\text{gen}}$, where the "gen" subscript indicates "generalized" moments including the auxiliary variables $\hat{\boldsymbol{x}}(t)$

$$\boldsymbol{\mu}_{\text{gen}}(t) = \begin{pmatrix} \langle \boldsymbol{x}(t) \rangle_Q \\ -\mathrm{i} \langle \hat{\boldsymbol{x}}(t) \rangle_Q \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}(t) \\ -\mathrm{i} \hat{\boldsymbol{\mu}}(t) \end{pmatrix} \tag{5.3.3}$$

$$\boldsymbol{C}_{\text{gen}}(t, t') = \left\langle \begin{pmatrix} \delta \boldsymbol{x}(t) \\ -\mathrm{i} \delta \hat{\boldsymbol{x}}(t) \end{pmatrix} \begin{pmatrix} \delta \boldsymbol{x}(t') & -\mathrm{i} \delta \hat{\boldsymbol{x}}(t') \end{pmatrix} \right\rangle_Q = \begin{pmatrix} \boldsymbol{C}(t, t') & \boldsymbol{R}(t, t') \\ \boldsymbol{R}(t', t)^T & \boldsymbol{B}(t, t') \end{pmatrix} \tag{5.3.4}$$

with $\delta \boldsymbol{x}(t) = \boldsymbol{x}(t) - \boldsymbol{\mu}(t)$ and $\delta \hat{\boldsymbol{x}}(t) = \hat{\boldsymbol{x}}(t) - \hat{\boldsymbol{\mu}}(t)$. For the sake of brevity, we will write $\langle \rangle_Q$ as simply $\langle \rangle$ and the notation used, similarly to chapters 2 and 3, has the following meaning

$$\begin{aligned} \boldsymbol{C}(t, t') =& \langle \delta \boldsymbol{x}(t) \delta \boldsymbol{x}^T(t') \rangle = \langle \boldsymbol{x}(t) \boldsymbol{x}^T(t') \rangle - \boldsymbol{\mu}(t) \boldsymbol{\mu}^T(t') \\ \boldsymbol{R}(t, t') =& -\mathrm{i} \langle \delta \boldsymbol{x}(t) \delta \hat{\boldsymbol{x}}^T(t') \rangle = -\mathrm{i} \langle \boldsymbol{x}(t) \hat{\boldsymbol{x}}^T(t') \rangle + \mathrm{i} \boldsymbol{\mu}(t) \hat{\boldsymbol{\mu}}^T(t') \\ \boldsymbol{B}(t, t') =& -\langle \delta \hat{\boldsymbol{x}}(t) \delta \hat{\boldsymbol{x}}^T(t') \rangle = -\langle \hat{\boldsymbol{x}}(t) \hat{\boldsymbol{x}}^T(t') \rangle + \hat{\boldsymbol{\mu}}(t) \hat{\boldsymbol{\mu}}^T(t') \end{aligned} \tag{5.3.5}$$

From general results for MSRJD path integrals [38, 75] it follows that $R_{ij}(t, t')$ has the meaning of a local response of $x_i(t)$ to a perturbing field $-\mathrm{i} \hat{x}_j(t')$ applied on some other node j at an earlier time t'; it is therefore non-vanishing only for $t > t'$. $B_{ij}(t, t')$, finally, is expected to vanish for all times t and t', as is $\hat{\mu}_i(t)$: we refer to [144] for a derivation of this property from the normalization requirement that the generating functional (5.2.11) at $\boldsymbol{\psi} = 0$ plays the role of a partition function.

The optimal Gaussian approximation can be found from the stationary points of the Kullback-Leibler (KL) Divergence between $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$ and $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ [114]

$$KL(Q||P) = \int D\boldsymbol{x}\, D\hat{\boldsymbol{x}}\, Q(\boldsymbol{x}, \hat{\boldsymbol{x}}) \ln \frac{Q(\boldsymbol{x}, \hat{\boldsymbol{x}})}{P(\boldsymbol{x}, \hat{\boldsymbol{x}})} \tag{5.3.6}$$

which are

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{\mu}_{\mathrm{gen}}} = 0 \tag{5.3.7a}$$

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{C}_{\mathrm{gen}}} = 0 \tag{5.3.7b}$$

The dynamics of the full probability distribution $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$, which would be intractable, is reduced to differential equations for the time-dependent parameters of the optimal Gaussian $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$, i.e. its first 2 moments. This calculation (explained in detail in the appendix G) basically consists of a reformulation in the MSRJD formalism of the variational derivation provided by Archambeau et al. in [110], whose main features are recovered. (In addition here we retain the nonlinear dependence on concentrations of the diffusion matrix $\boldsymbol{\Sigma}(\boldsymbol{x}(t))$). We remark that the positivity and the existence of a minimum for the KL are not fully guaranteed in the case we consider, as by construction $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$ and $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ can have complex values. The KL minimization (mathematically rigorous only for real valued distributions) is thus applied as a *heuristic* strategy to find the optimal approximation. This choice can be *a posteriori* validated by comparison with the known results in [110].

The evolution of means and correlations shows that the GVA trajectories are effectively those of the Langevin dynamics (5.2.1) linearized around the time-dependent means (thus valid locally in their vicinity). One has

$$\frac{d\mu_i(t)}{dt} = \langle \Phi_i(\boldsymbol{x}(t)) \rangle \tag{5.3.8}$$

$\boldsymbol{\mu}(t)$ evolves under the average drift of the initial equation of motion (5.2.1), still retaining its non-linearities in the form of a combination of means and correlations. Equation (5.3.8) sets the values of the mean of a nonlinear theory approximated by a Gaussian; once this is given, equation (5.3.8) is equivalent to the stationary path approximation [75] as for a probability density functional of Gaussian form, which is exact in a dynamical linear theory[2], the average path is identical to the stationary one, given certain initial conditions. Dynamical expressions for auxiliary variables are compatible with $\hat{\boldsymbol{\mu}}(t) \equiv 0\ \forall t$, as prescribed by the normalization condition, and $\boldsymbol{R}(t, t) \equiv 0\ \forall t$ (vanishing equal times responses are a generic consequence of the Itô discretization [23, 71]).

[2]We recall that for a linear dynamics, as exploited also in chapter 2 and 3, the path distribution is a bilinear form of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, thus a Gaussian.

Furthermore, the equation for the auxiliary variables correlator is compatible with $\boldsymbol{B}(t,t') \equiv 0$ $\forall t', t$, as expected. This effective linearization corresponds to the vectorial equation

$$\frac{d(\boldsymbol{x}(t)-\boldsymbol{\mu}(t))}{dt} = \boldsymbol{K}(t)(\boldsymbol{x}(t)-\boldsymbol{\mu}(t)) + \boldsymbol{\xi}(t) \tag{5.3.9}$$

with an effectively linear rate matrix $\boldsymbol{K}(t)$ defined as

$$K_{ij}(t) = \left\langle \frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \right\rangle \tag{5.3.10}$$

and $\boldsymbol{\xi}(t)$ effectively distributed as

$$\langle \boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t')\rangle = \langle \boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \delta(t-t') \tag{5.3.11}$$

so the noise (diffusion) covariance in the Gaussian picture is the true one averaged over the Gaussian approximating distribution. The result (5.3.10) coincides with GVA already developed by Archambeau et al. [110] by KL divergence minimization, while (5.3.11) generalizes that approach to the case in which there is a mismatch between the diffusion matrices of the true and the approximated process. Responses and correlations obtained by the GVA are the ones one could straightforwardly derive from (5.3.9) (see also appendix G). Namely

$$\boldsymbol{R}(t',t) = \theta(t'-t)\,\mathrm{T}\left[e^{\int_t^{t'} \boldsymbol{K}(t'')dt''}\right] \tag{5.3.12}$$

similarly to what has been derived in chapter 2, section 2.4.3. The time-ordering operator T has been here introduced because time is used as an index to determine the order of $\boldsymbol{K}(t)$ matrices: earlier times on the right and later times on the left. The correlations are connected to response functions by

$$\boldsymbol{C}(t',t) = \boldsymbol{R}(t',0)\,\boldsymbol{C}(0,0)\,\boldsymbol{R}^T(t,0) + \int_0^{\min(t,t')} \boldsymbol{R}(t',s)\,\langle \boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle\,\boldsymbol{R}^T(t,s)ds \tag{5.3.13}$$

and obey the differential equation

$$\frac{\partial \boldsymbol{C}(t',t)}{\partial t} = \boldsymbol{C}(t',t)\boldsymbol{K}^T(t) + \boldsymbol{R}(t',t)\langle \boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \tag{5.3.14}$$

The first term, containing the noise covariance $\boldsymbol{\Sigma}$, is the *source* term and it tends to increase the amplitude of fluctuations, while the second one, which comprises the rate matrix $\boldsymbol{K}$, is called *propagation* term and it drives fluctuations to decay exponentially. For equal times one has

$$\frac{\partial \boldsymbol{C}(t,t)}{\partial t} = \boldsymbol{C}(t,t)\boldsymbol{K}^T(t) + \boldsymbol{K}(t)\boldsymbol{C}(t,t) + \langle \boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \tag{5.3.15}$$

from which, by looking at the steady state, thus by setting $\partial_t \boldsymbol{C}(t,t) = 0$, the Lyapunov equation can be straightforwardly read

$$\boldsymbol{C}\boldsymbol{K}^T + \boldsymbol{K}\boldsymbol{C} + \boldsymbol{\Sigma} = 0 \tag{5.3.16}$$

for the stationary $\boldsymbol{C} = \boldsymbol{C}(0) = \boldsymbol{C}(t-t)$, where $\boldsymbol{K}$ and $\boldsymbol{\Sigma}$ are constant as calculated at the steady state values.

We shall stress that the GVA differs from the Linear Noise Approximation (LNA)[3], whose solution is a Gaussian distribution but corresponding to a linearization of the CLE around the macroscopic (deterministic) trajectory (see [64]). For this one would have

$$\frac{d\mu_i(t)}{dt} = \Phi_i(\boldsymbol{\mu}(t)) \tag{5.3.17}$$

while in the GVA, the mean trajectory is given by (5.3.8) where $\langle \Phi_i(\boldsymbol{x}(t)) \rangle = \Phi_i(\boldsymbol{\mu}(t), \boldsymbol{C}(t,t))$. The mean dynamics (5.3.8), and consequently also the effective drift (5.3.10) and the effective diffusion matrix (5.3.11), contain as well equal time correlations, i.e. terms stemming from fluctuations which can be the regarded as finite volume corrections. Such corrections tend to vanish for large volumes as this is the behaviour of source term (5.2.4), i.e. $\Phi_i(\boldsymbol{\mu}(t), \boldsymbol{C}(t,t)) \sim \Phi_i(\boldsymbol{\mu}(t))$, thus only in this limit the GVA $\boldsymbol{\mu}(t)$ coincides with the deterministic (macroscopic) solution of the initial equation (5.2.1) and becomes equivalent to the LNA.

5.4 Subnetwork effective description within GVA

We shall exploit the GVA to find an effective description for the species in the subnetwork accounting also for the presence of the external bulk: this constitutes the novel contribution of this work.

5.4.1 Bulk and subnetwork

The first step is to simplify variational equations with the assumed bulk/subnetwork separation, that need to be reflected in the covariance matrix structure. We thus introduce a block structure for $\boldsymbol{C}$ and $\boldsymbol{R}$

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}^{\mathrm{ss}} & \boldsymbol{C}^{\mathrm{sb}} \\ \boldsymbol{C}^{\mathrm{bs}} & \boldsymbol{C}^{\mathrm{bb}} \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}^{\mathrm{ss}} & \boldsymbol{R}^{\mathrm{sb}} \\ \boldsymbol{R}^{\mathrm{bs}} & \boldsymbol{R}^{\mathrm{bb}} \end{pmatrix}$$

where the superscript (s or b) indicates which subset (subnetwork or bulk) the relevant variable belongs to. One could look at restricted approximations where we suppose some decorrelation

[3]The LNA is obtained from the second order terms in $1/\sqrt{V}$ of the Van Kampen system size expansion of the Master Equation [10], whereby the first order gives the macroscopic rate equations.

between subnetwork or bulk, yet the key goal is to find a description just for the subnetwork while retaining the feedbacks arising via the bulk. Terms that keep track of this exchange (e.g. $\boldsymbol{C}^{\mathrm{bs}}_{\mathrm{gen}}$, $\boldsymbol{C}^{\mathrm{sb}}_{\mathrm{gen}}$) should reasonably be present. If the off-diagonal blocks $\boldsymbol{C}^{\mathrm{sb}}_{\mathrm{gen}}, \boldsymbol{C}^{\mathrm{sb}}_{\mathrm{gen}} \equiv 0$, we would obtain a description in which bulk and subnetwork evolve separately, thus these blocks are fundamental to understand how the bulk drives subnetwork dynamics (this can be better appreciated by looking at the derivation of the reduced subnetwork dynamics we provide in appendix). For example the ansatz of a factorized approximation (i.e. in which the joint probability over species factorizes into the product of distributions of single species) would lead to vanishing $\boldsymbol{C}^{\mathrm{sb}}_{\mathrm{gen}}, \boldsymbol{C}^{\mathrm{sb}}_{\mathrm{gen}}$, thus would prevent one from capturing bulk feedbacks into the subnetwork dynamics. Such feedbacks will appear in the evolution of subnetwork in the form of a memory, as expected from projection methods, which establish the structure of the reduced dynamics.

Marginalization

Memory terms from the Gaussian-process representation of the dynamics can be obtained via bulk marginalization. Practically the idea is to read off from the variational approximation what the marginal measure over subnetwork trajectories is. In fact, once we have $\boldsymbol{C}^{\mathrm{ss}}$ from the Gaussian variational approximation (allowing for full correlations between subnetwork and bulk) we can ask: *what equations of motion does this Gaussian distribution over subnetworks encode?* As we mentioned, they are expected to include a memory term and a coloured noise. In principle, we can retrieve them by finding the (operator) inverse $(\boldsymbol{C}^{\mathrm{ss}})^{-1}$ to give the marginal dynamical action for subnetwork. Nevertheless, starting from a dynamics linearized around time-dependent means one can proceed straightforwardly by substitution, as follows.

Linearizing around the time-dependent subnetwork and bulk means $\boldsymbol{\mu}^{\mathrm{s}}(t)$ and $\boldsymbol{\mu}^{\mathrm{b}}(t)$, single concentrations can be rewritten as $\boldsymbol{x}^{\mathrm{s}}(t) = \boldsymbol{\mu}^{\mathrm{s}}(t) + \delta\boldsymbol{x}^{\mathrm{s}}(t)$ and $\boldsymbol{x}^{\mathrm{b}}(t) = \boldsymbol{\mu}^{\mathrm{b}}(t) + \delta\boldsymbol{x}^{\mathrm{b}}(t)$. We can apply to the rate matrix in equation (5.3.9) the subnetwork/ bulk blocks decomposition

$$\boldsymbol{K}(t) = \begin{pmatrix} \boldsymbol{K}^{\mathrm{ss}}(t) & \boldsymbol{K}^{\mathrm{sb}}(t) \\ \boldsymbol{K}^{\mathrm{bs}}(t) & \boldsymbol{K}^{\mathrm{bb}}(t) \end{pmatrix}$$

We switch to concentrations deviations from the means $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ and $\delta\boldsymbol{x}^{\mathrm{b}}(t)$ obeying the Langevin dynamics

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}}{dt} = \boldsymbol{K}^{\mathrm{ss}}(t)\delta\boldsymbol{x}^{\mathrm{s}} + \boldsymbol{K}^{\mathrm{sb}}(t)\delta\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{\xi}^{\mathrm{s}} + \delta\boldsymbol{x}^{\mathrm{s}}(0)\delta(t) \tag{5.4.1a}$$

$$\frac{d\delta\boldsymbol{x}^{\mathrm{b}}}{dt} = \boldsymbol{K}^{\mathrm{bb}}(t)\delta\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{K}^{\mathrm{bs}}(t)\delta\boldsymbol{x}^{\mathrm{s}} + \boldsymbol{\xi}^{\mathrm{b}} + \delta\boldsymbol{x}^{\mathrm{b}}(0)\delta(t) \tag{5.4.1b}$$

where the last terms impose the initial conditions $\delta\boldsymbol{x}^{\mathrm{s}}(0)$ and $\delta\boldsymbol{x}^{\mathrm{b}}(0)$. The reduced dynamics for the single subnetwork deviations from the means $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ should read

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} = \boldsymbol{K}^{\mathrm{ss}}(t)\delta\boldsymbol{x}^{\mathrm{s}}(t) + \int_0^t dt'(\boldsymbol{M}^{\mathrm{ss}})^T(t,t')\delta\boldsymbol{x}^{\mathrm{s}}(t') + \boldsymbol{\chi}(t) \tag{5.4.2}$$

All subnetwork reactions contribute directly via local-in-time term $\boldsymbol{K}^{\mathrm{ss}}$ which contains them in the original form, as if the subnetwork was isolated, while the bulk gives a contribution via the additional non-local-in-time terms. The second one contains the memory function, denoted as $\boldsymbol{M}^{\mathrm{ss}}$, while $\boldsymbol{\chi}$ stands for the effective noise in the reduced system. Memory and noise terms of the effective subnetwork description can at this point be calculated by directly eliminating $\delta\boldsymbol{x}^{\mathrm{b}}$, i.e. we can solve (5.4.1b) for the bulk as a function of subnetwork variables

$$\delta\boldsymbol{x}^{\mathrm{b}}(t) = \mathrm{T}\left[e^{\int_0^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\delta\boldsymbol{x}^{\mathrm{b}}(0) + \boldsymbol{K}^{\mathrm{sb}}(t)\int_0^t \mathrm{T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\left(\boldsymbol{K}^{\mathrm{bs}}(t')\delta\boldsymbol{x}^{\mathrm{s}}(t') + \boldsymbol{\xi}^{\mathrm{b}}(t')\right)dt' \tag{5.4.3}$$

and next we substitute into the equation for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ (5.4.1a). One can then read the explicit expression for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ effective dynamics, whose structure is symbolically (5.4.2). For example, the memory function is given by

$$(\boldsymbol{M}^{\mathrm{ss}})^T(t,t') = \boldsymbol{K}^{\mathrm{sb}}(t)\,\mathrm{T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\boldsymbol{K}^{\mathrm{bs}}(t') \tag{5.4.4}$$

We emphasize an essential property of the memory, which is clear from (5.4.4), the *boundary* structure: if we define "boundary species" as the ones interacting with the bulk (thus containing some coefficient of $\boldsymbol{K}^{\mathrm{sb}}$), only the equations for these contain memory terms. For the effective noise covariance, we have

$$\begin{aligned}\boldsymbol{N}_0^{\mathrm{ss}}(t,t') &= \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{K}^{\mathrm{sb}}(t)\boldsymbol{R}^{\mathrm{bb}}(t,0)\boldsymbol{C}^{\mathrm{bb}}(0,0)\boldsymbol{R}^{\mathrm{bb}\,T}(t',0)(\boldsymbol{K}^{\mathrm{sb}})^T(t') \\ &+ \langle\boldsymbol{\Sigma}^{\mathrm{ss}}(\boldsymbol{x}(t))\rangle\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}(t)\int_0^{\min(t,t')} dt''\boldsymbol{R}^{\mathrm{bb}}(t,t'')\langle\boldsymbol{\Sigma}^{\mathrm{bb}}(\boldsymbol{x}(t))\rangle\boldsymbol{R}^{\mathrm{bb}\,T}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T(t'')\end{aligned} \tag{5.4.5}$$

where we have defined the response as $\boldsymbol{R}^{\mathrm{bb}}(t,t'') = \theta(t-t'')\mathrm{T}\left[e^{\int_{t''}^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]$ and for simplicity we have considered the noise and the initial conditions in bulk and subnetwork reciprocally independent (i.e. $\boldsymbol{C}^{\mathrm{sb}}(0,0) = \boldsymbol{C}^{\mathrm{bs}}(0,0) = \boldsymbol{\Sigma}^{\mathrm{sb}} = \boldsymbol{\Sigma}^{\mathrm{bs}} = \mathbb{0}$). We note that for a linear stochastic dynamics, i.e. a purely Gaussian process, memory and effective noise can be derived also by simple matrix inversion, as explained in appendix H. This simple derivation by direct substitution of the bulk solution clarifies conceptually how non-Markovian terms in the subnetwork dynamics emerge via the dynamical modulation of the bulk.

Although very interesting in order to understand the logic behind model reduction, results (5.4.4) and (5.4.5) are not convenient computationally because of the integrals implied by the time dependence of $\boldsymbol{K}^{\mathrm{bb}}(t)$. For the sake of computational and conceptual simplicity, we shall consider

thus, from now on, a dynamics linearized around the *steady states*: this implies that the diffusion matrix and the dynamical matrix are both evaluated at steady state, i.e. they are constant in time $\langle \boldsymbol{\Sigma}(\boldsymbol{x}(t)) \rangle \equiv \boldsymbol{\Sigma}$ and $\boldsymbol{K}(t) \equiv \boldsymbol{K}$. In this simpler case (5.4.4) becomes

$$(\boldsymbol{M}^{\mathrm{ss}})^T(t,t') = \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} \boldsymbol{K}^{\mathrm{bs}} \tag{5.4.6}$$

In addition, the linearization around steady states allows a comparison with the reduced description derived via projection methods by Rubin et al. [77] (and summarized in section J.2 of the appendix J). Its validity is limited to the vicinity of steady states and a generalization to time-dependent means is not straightforward as in the GVA. The expression for the memory function (5.4.6) fully corresponds to the one in [77]. We refer then to [77] for a systematic analysis of typical amplitudes and timescales of the memory, which encode its overall effect on dynamics, in the case of a mass action kinetics as (5.2.1). For the noise covariance one has

$$\begin{aligned} \boldsymbol{N}_0^{\mathrm{ss}}(t,t') =& \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{\Sigma}^{\mathrm{ss}}\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{R}^{\mathrm{bb}}(t,0)\boldsymbol{C}^{\mathrm{bb}}(0,0)\boldsymbol{R}^{\mathrm{bb}\,T}(t',0)(\boldsymbol{K}^{\mathrm{sb}})^T \\ &+ \boldsymbol{K}^{\mathrm{sb}} \int_0^{\min(t,t')} dt'' \boldsymbol{R}^{\mathrm{bb}}(t,t'')\boldsymbol{\Sigma}^{\mathrm{bb}}\boldsymbol{R}^{\mathrm{bb}\,T}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T \end{aligned} \tag{5.4.7}$$

with the bulk response now simply given by $\boldsymbol{R}^{\mathrm{bb}}(t,t'') = \theta(t-t'')e^{\boldsymbol{K}^{\mathrm{bb}}(t-t'')}$. This noise term is comparable but does not agree completely with the random force in the projection approach, because of differences in some fundamental assumptions of the two methods (see appendix J). One sees that $\boldsymbol{N}_0^{\mathrm{ss}}$ is composed of two contributions, one connected to the white noise in the bulk dynamics and the other to the initial bulk uncertainty: only the latter basically resembles the random force correlator of projection methods (see section J.2). Therefore, in a linearized dynamics, the predictions of the variational approach agree with the ones of the projection framework in the large volume limit (for which the intrinsic fluctuations vanish and only the initial conditions contribution is left).

So far no assumption on the relative timescales of bulk and subnetwork has been made. If we restrict to the case of a slowly varying subnetwork embedded in a fast bulk, equation (5.4.2) rgduces to the slow-scale LNA introduced by Thomas et al. in [4, 131] in the large volume limit: we regard it as an important consistency check for our approach. In fact, let us write it explicitly taking as initial time $t = -\infty$

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} = \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) + \int_{-\infty}^{t} dt' \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t') + \boldsymbol{\xi}^{\mathrm{s}}(t) + \int_{-\infty}^{t} dt' \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{\xi}^{\mathrm{b}}(t') \tag{5.4.8}$$

Here we have used (5.4.6) for the memory kernel and we have derived the noise with covariance (5.4.7) with both subnetwork and bulk initially at steady state, i.e.

$$\boldsymbol{\chi}(t) = \boldsymbol{\xi}^{\mathrm{s}}(t) + \int_{-\infty}^{t} dt' \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{\xi}^{\mathrm{b}}(t') \tag{5.4.9}$$

With large volumes, $\boldsymbol{K}$ gives the effective drift of a linearization around deterministic steady states (i.e. the GVA corresponds to the LNA). By assuming the subnetwork as slowly varying, we can substitute $\delta\boldsymbol{x}^{\mathrm{s}}(t') \sim \delta\boldsymbol{x}^{\mathrm{s}}(t)$ in the memory integral of of (5.4.8). In addition, as the bulk fluctuations are fast, the ones which affect the dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}$ at time t are the ones temporally very close, thus we can write $\boldsymbol{\xi}^{\mathrm{b}}(t') \sim \boldsymbol{\xi}^{\mathrm{b}}(t)$ in the effective noise integral (5.4.9): this equivalent to state that for large $\boldsymbol{K}^{\mathrm{bb}}$ the coloured noise (5.4.9) behaves as a white noise. (This replacement is heuristic, as the noise is not well defined as a function, but one can obtain the same result by properly looking at the correlator, whose temporal part reduces to a δ-function in this limit). As a consequence in both time integrals on the r.h.s. of (5.4.8) we are left with $\int_{-\infty}^{t} dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} = -\boldsymbol{K}^{\mathrm{bb}\,-1}$. The equation resulting from this approximation is exactly the slow-scale LNA Langevin equation (expression 38 of [131]), which in our notation can be cast as

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} = (\boldsymbol{K}^{\mathrm{ss}} - \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{K}^{\mathrm{bb}\,-1}\boldsymbol{K}^{\mathrm{bs}})\delta\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{\xi}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{K}^{\mathrm{bb}\,-1}\boldsymbol{\xi}^{\mathrm{b}}(t) \tag{5.4.10}$$

5.5 Nonlinear corrections by perturbation theory

5.5.1 Nonlinear memory

More generally, to describe the dynamics of the subnetwork we need to integrate out bulk variables to obtain an "effective" action (where by effective we mean depending *solely* on the subnetwork), corresponding to a marginal probability distribution; explicitly it can be written

$$\mathcal{H}_{\mathrm{eff}}(\boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}) = \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{\mathcal{H}(\boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}})} \tag{5.5.1}$$

where $D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}}$ indicate the integration over bulk paths. To go beyond the simple linear expression for the memory function and the effective coloured noise of section 5.4.1, one can develop a perturbative expansion for the nonlinear contributions in the dynamics, giving cubic terms in the effective action of the subnetwork. According to perturbation theory, one decomposes the action $\mathcal{H}$ into a "non-interacting" part $\mathcal{H}_0$ (containing only the purely Gaussian terms, i.e. quadratic in all variables) and an interacting one $\Delta\mathcal{H}$ (containing higher powers). The latter can be considered, formally, as a *perturbation*, provided that the couplings by which they enter the dynamics are small. The expression (5.5.1), by writing $\mathcal{H} = \mathcal{H}_0 + \Delta\mathcal{H}$, becomes

$$\begin{aligned} \mathcal{H}_{\mathrm{eff}} &= \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{(\mathcal{H}_0 + \Delta\mathcal{H})} = \\ &= \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{\mathcal{H}_0} \left(\mathbb{1} + \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} Q_0(\boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}} | \boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}) \Delta\mathcal{H} + O(\Delta\mathcal{H})^2 + \ldots \right) \end{aligned} \tag{5.5.2}$$

with

$$-\Delta\mathcal{H} = \int_0^T dt \left\{ \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,\mathrm{T}} \Big[\boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}) + \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}) + \boldsymbol{K}^{\mathrm{s,bb}}(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}) + \right.$$
$$\left. \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,\mathrm{T}} \Big[\boldsymbol{K}^{\mathrm{b,ss}}(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}) + \boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}) + \boldsymbol{K}^{\mathrm{b,bb}}(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}) \Big] \right\} \quad (5.5.3)$$

where $\boldsymbol{K}^{\mathrm{s,ss}}$, $\boldsymbol{K}^{\mathrm{s,sb}}$, $\boldsymbol{K}^{\mathrm{b,sb}}$, $\boldsymbol{K}^{\mathrm{b,bb}}$ are used as shorthands for the set of rate constants involved, respectively, in the reaction of two subnetwork species ($\boldsymbol{K}^{\mathrm{s,ss}}$) giving another subnetwork species or a bulk and subnetwork species giving a complex in the subnetwork ($\boldsymbol{K}^{\mathrm{s,sb}}$) or in the bulk ($\boldsymbol{K}^{\mathrm{b,sb}}$) or two bulk species ($\boldsymbol{K}^{\mathrm{b,bb}}$) giving another bulk species. We will set instead $\boldsymbol{K}^{\mathrm{s,bb}} = \boldsymbol{K}^{\mathrm{b,ss}} \equiv 0$, i.e. we do not include in the subnetwork complexes formed by 2 bulk species (and viceversa we do not include in the bulk complexes formed by 2 subnetwork species). Another biologically sensible assumption, capable of simplifying even more expression (5.5.3), would be that a subnetwork complex can be created only by two subnetwork species (i.e. $\boldsymbol{K}^{\mathrm{s,sb}} \equiv 0$), whereas a bulk complex can be created either by two bulk species or a bulk and subnetwork protein. We are in fact interested in small subnetworks, which in principle contain only species well characterized quantitatively; following this logic, complexes in the subnetwork are the ones whose formation can be wholly tracked, while complexes with under-resolved time courses are assigned to the bulk. (We will implement this assumption in the application of our method to biochemical networks of section 5.6). We use the $\boldsymbol{a} \circ \boldsymbol{b}$ notation, not to be confused with a Hadamard (elementwise) product, to denote the outer product $\boldsymbol{a}\boldsymbol{b}^T$ rearranged into a single (column) vector. This vector then has as its entries all possible componentwise products $a_i b_j$ so that e.g. $\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}$ is a vector of dimension $N^{\mathrm{s}}N^{\mathrm{b}}$. The matrices $\boldsymbol{K}^{\mathrm{s,ss}}$ etc. are then defined to contain the appropriate coefficients so that, for example,

$$[\boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}})]_i = \sum_{j,l=1}^{N^{\mathrm{s}}} \left(\frac{1}{2} k^+_{jl,i} \delta x_j \delta x_l - k^+_{ij,l} \delta x_i \delta x_j \right) + \frac{1}{2} \sum_{j=1}^{N^{\mathrm{s}}} k^+_{jj,i} \delta x_j \delta x_j - \sum_{l=1}^{N^{\mathrm{s}}} k^+_{ii,l} \delta x_i \delta x_i \quad (5.5.4)$$

as can be seen from (5.2.2).

The reduced action should depend only on subnetwork variables *by definition*: the basic idea is then to fix subnetwork variables and express bulk ones conditionally on them. To estimate the perturbative corrections we therefore need bulk *conditional* Gaussian moments, which, using the Wick's theorem, can be decomposed into combinations of conditional correlations and means (this procedure is explained in more detail in appendix I). At this point a choice of the Gaussian conditional distribution Q_0 should be made. As a natural option, we could keep the GVA approximating distribution, thus set $Q_0 = Q$, but this would lead to additional "field" type terms (i.e. not dependent neither on $\delta\boldsymbol{x}^{\mathrm{s}}$ nor $\delta\boldsymbol{x}^{\mathrm{b}}$) in the perturbative action which should be treated separately. To see

it, let us take a schematic version of the full action

$$-\mathcal{H} = \int_0^T dt \sum_i \mathrm{i}\hat{x}_i \, \Phi_i(\boldsymbol{x}) = \int_0^T dt \sum_i \mathrm{i}\hat{x}_i \, \Phi_i(\boldsymbol{\mu} + \delta\boldsymbol{x}) \sim \int_0^T dt \sum_i \mathrm{i}\hat{x}_i \, [\Phi_i(\boldsymbol{\mu}) + O(\delta\boldsymbol{x}) + O(\delta\boldsymbol{x}^2)] \tag{5.5.5}$$

where in the last step we have performed an expansion of the drift Φ_i around steady states means $\boldsymbol{\mu}$ up to quadratic terms in $\delta\boldsymbol{x}$. "Field" terms would appear as in the GVA $\Phi_i(\boldsymbol{\mu})$ is not zero: it is therefore more convenient to appeal to the LNA [64], for which $\Phi_i(\boldsymbol{\mu}) = 0$. Note that given this choice the role of $\boldsymbol{\mu}$ here changes, i.e. we are interpreting it not anymore as the steady states of the *mean* dynamics but as the ones of the *deterministic* dynamics (see also the explanation in section 5.3). The resulting quadratic action is

$$\begin{aligned}\mathcal{H}_0 \;=\; \int_0^T dt \Big\{ & \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)[\partial_t \delta\boldsymbol{x}^{\mathrm{s}} - \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}}\delta\boldsymbol{x}^{\mathrm{b}}(t)] + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{ss}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{sb}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) \\ & \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,T}(t)[\partial_t \delta\boldsymbol{x}^{\mathrm{b}} - \boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{bb}}\delta\boldsymbol{x}^{\mathrm{b}}(t)] + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{bb}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{bs}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) \Big\}\end{aligned} \tag{5.5.6}$$

General formulas for conditioning Gaussian quantities (see [16] and section I.1 of appendix I) give for auxiliary variables

$$\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t) = \int_t^T dt' e^{-(\boldsymbol{K}^{\mathrm{bb}})^T(t-t')}(\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{5.5.7}$$

with the boundary condition

$$\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(T) = 0 \tag{5.5.8}$$

which can be interpreted as an "initial" condition for a temporally reversed evolution. The dynamics for the auxiliary variables consists thus of a temporal backward propagation and the boundary condition for $\hat{\boldsymbol{\mu}}(T)$ must be specified rather than value at $t = 0$. This fact is consistent with the theory of conditional Markov processes, for which calculating posterior distributions requires information to propagate both in the forward and backward direction (see also chapter 3). Whereas marginal auxiliary means are expected to vanish for consistency of the formalism [23, 144], the conditional ones do not, as long as the subnetwork values one is conditioning on are not identically zero (in (5.5.7) we have actually used that the marginal mean $\hat{\boldsymbol{\mu}}^{\mathrm{s}}, \hat{\boldsymbol{\mu}}^{\mathrm{b}} \equiv 0$). The conditional bulk mean concentration reads as follows

$$\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t) = \boldsymbol{\mu}^{\mathrm{b}} + \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt' \boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t')(\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{5.5.9}$$

and incorporates separately a term with $\delta\boldsymbol{x}^{\mathrm{s}}$ and one with $\hat{\boldsymbol{x}}^{\mathrm{s}}$. For the sake of brevity we call them $\boldsymbol{\nu}(t) = \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')$ and $\hat{\boldsymbol{\nu}}(t) = -\int_0^T dt' \boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t')(\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t')$, so that we can write $\delta\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t) = \boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t) - \boldsymbol{\mu}^{\mathrm{b}} = \boldsymbol{\nu}(t) + \hat{\boldsymbol{\nu}}(t)$. While $\boldsymbol{\nu}(t)$ is the deterministic bulk solution, $\hat{\boldsymbol{\nu}}(t)$ carries the "stochastic" contributions from the bulk, i.e. the uncertainty on its initial values and its intrinsic

noise. The initial condition for $\boldsymbol{\mu}^{\mathrm{b|s}}(t)$ is connected to $\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0)$

$$\boldsymbol{\mu}^{\mathrm{b|s}}(0) = -\boldsymbol{C}^{\mathrm{bb}}(0,0)\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0) \tag{5.5.10}$$

as derived in appendix I (see section I.1). The conditional second moments are the ones describing the bulk as if isolated from the subnetwork. As such, the conditional correlator of auxiliary variables $\boldsymbol{B}^{\mathrm{bb|s}}(t,t'') \equiv 0$ and the equal time conditional response $\boldsymbol{R}^{\mathrm{bb|s}}(t,t) \equiv 0$, analogously to the marginal ones. Finally the conditional correlator is given by

$$\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') = \int_0^{\min(t,t'')} dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{\Sigma}^{\mathrm{bb}}e^{(\boldsymbol{K}^{\mathrm{bb}})^T(t''-t')} + e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)e^{(\boldsymbol{K}^{\mathrm{bb}})^T t''} \tag{5.5.11}$$

To derive the reduced subnetwork action, what is needed is to make explicit the dependence on the subnetwork: while the conditional mean (5.5.9) is a linear function of $\delta\boldsymbol{x}^{\mathrm{s}}$, the second order statistics (5.5.11) does not depend on this variable. It is in fact a general property of Gaussian processes that the conditional variance does not depend on the way of conditioning (i.e. on the particular values of the variables being conditioned on). By inserting (5.5.11), (5.5.9) and (5.5.7) in (5.5.2), one obtains the reduced action as function of *solely* subnetwork degrees of freedom; from it, one can read by inspection the nonlinear reduced dynamics of the subnetwork

$$\begin{aligned}\frac{d\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} &= \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t)) + \int_0^t \boldsymbol{M}^{\mathrm{ss}\,T}(t,t')\delta\boldsymbol{x}^{\mathrm{s}}(t')dt' \\ &+ \int_0^t dt'\int_0^t dt''\boldsymbol{M}^{\mathrm{s,ss}\,T}(t,t'',t')(\delta\boldsymbol{x}^{\mathrm{s}}(t'')\circ\delta\boldsymbol{x}^{\mathrm{s}}(t')) + \boldsymbol{\chi}(t)\end{aligned} \tag{5.5.12}$$

where $\boldsymbol{\chi}(t)$, the coloured noise, has a covariance $\langle\boldsymbol{\chi}(t)\boldsymbol{\chi}(t'')^T\rangle = \boldsymbol{N}_0^{\mathrm{ss}}(t,t'') + \boldsymbol{N}_1^{\mathrm{ss}}(t,t'')$ consisting of a linear and nonlinear contribution. While $\boldsymbol{N}_0^{\mathrm{ss}}(t,t'')$ is the covariance of a Gaussian coloured noise (namely (5.4.7)), $\boldsymbol{N}_1^{\mathrm{ss}}(t,t'')$ is non-Gaussian as linearly dependent on $\boldsymbol{x}^{\mathrm{s}}$ (see section I.2 of appendix I).

We see that reactions within the subnetwork contribute to the dynamical equations for subnetwork only via $\boldsymbol{K}^{\mathrm{ss}}$ and $\boldsymbol{K}^{\mathrm{s,ss}}$ (a compact notation for the rate constants of respectively linear and nonlinear couplings internal to the subnetwork): so all subnetwork reactions are captured, in their original form, in local-in-time terms, one of the *desiderata* of our coarse-grained description. Importantly this implies also that subnetwork dynamics could be treated without initial approximations, e.g. the linearization around steady states as in this case, which could be applied only to boundary and bulk-bulk reactions.

The nonlinear corrections to the memory, $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')$, then correspond to the coefficients of terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$ in the effective perturbative action (5.5.2), while the ones to the effective noise covariance, $\boldsymbol{N}_1^{\mathrm{ss}}(t,t'')$, are extracted from terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$. More specifically, the first ones come

from the $\boldsymbol{\nu}(t)$ part of the conditional mean (5.5.9), while the latter from $\hat{\boldsymbol{\nu}}(t)$; we refer to section I.2 of appendix I for the steps needed in this regard and here we just state the final result (for $t > t'' > t'$)

$$\begin{aligned}\boldsymbol{M}^{\mathrm{s,ss}\,T}(t,t'',t')(\delta\boldsymbol{x}^{\mathrm{s}}(t'')\circ\delta\boldsymbol{x}^{\mathrm{s}}(t')) = \\ &+ \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'))\delta(t''-t) \\ &+ \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t'')}\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t'')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t''-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')) \\ &+ 2\int_{t''}^{t} ds\, \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-s)}\boldsymbol{K}^{\mathrm{b,bb}}(e^{\boldsymbol{K}^{\mathrm{bb}}(s-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'))\end{aligned} \tag{5.5.13}$$

5.5.2 Numerical implementation

The total memory acting on the dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ (the vector $\boldsymbol{\mathcal{M}}(t)$) is given by a time convolution modulated by the functions (5.4.6) and (5.5.13)

$$\boldsymbol{\mathcal{M}}(t) = \int_0^t \boldsymbol{M}^{\mathrm{ss}\,T}(t,t')\delta\boldsymbol{x}^{\mathrm{s}}(t')dt' + \int_0^t dt' \int_0^t dt''\boldsymbol{M}^{\mathrm{s,ss}\,T}(t,t'',t')(\delta\boldsymbol{x}^{\mathrm{s}}(t'')\circ\delta\boldsymbol{x}^{\mathrm{s}}(t')) \tag{5.5.14}$$

It consists of terms involving a different number of time integrals: each one can be represented via the solutions of additional differential equations. One can think of solving *integro-differential* equations (5.5.12) simply by means of a *differential* equations solver in an enlarged space of variables. We shall first rewrite (5.5.14) in terms of the $\boldsymbol{\nu}$ part of bulk conditional means

$$\begin{aligned}\boldsymbol{\mathcal{M}}(t) =&\boldsymbol{K}^{\mathrm{sb}}\boldsymbol{\nu}(t) + \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{\nu}(t))+ \\ &\int_0^t dt'\,\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\left[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{\nu}(t')) + \boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t'))\right]\end{aligned} \tag{5.5.15}$$

By comparing (5.5.14) (with the substitution of (5.4.6) and (5.5.13)) and (5.5.15), we see that some time integrals can be equally accounted for via the N^{b} differential equations for $\boldsymbol{\nu}(t)$

$$\frac{d}{dt}\boldsymbol{\nu} = \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{\nu} + \boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}} \tag{5.5.16}$$

The $\boldsymbol{K}^{\mathrm{b,sb}}$ and the $\boldsymbol{K}^{\mathrm{b,bb}}$ pieces of (5.5.15) contain an additional time integral; to translate them into differential equations for additional variables we need to apply a decomposition into eigenvalues and eigenvectors of $\boldsymbol{K}^{\mathrm{bb}}$ following the procedure provided in [135]. We briefly reformulate this implementation route for the GVA in appendix K, where we show in more detail that solving the N^{s} subnetwork equations (5.5.12) with integral memory terms is equivalent to solving a system with $2N^{\mathrm{b}}$ additional equations.

For projection methods, the additional variables one needs to introduce for expressing memory integrals via differential equations (see [77]) increases *quadratically* in the size of the bulk while

in the GVA only *linearly*. Importantly, such a difference can translate into a significant advantage from the point of view of numerical computations when a bulk of large size is taken into account. This is a plausible setting of application for these two model reduction approaches, as subnetworks of well resolved species realistically comprise just a few of them.

5.5.3 Comparison with projection methods

The comparison of GVA to projection methods when the full nonlinear dynamics is taken into account is not straightforward as for the linearized equations. The reason is that memory functions and the coloured noise correlators derived in the two approaches are not the same, if taken *separately*. For example, let us focus on the memory term in (5.5.12): both subnetwork species in the integral must be boundary ones, thus it describes how the past evolution of subnetwork boundary species might affect the present value of another subnetwork boundary species. This does not hold true for the projected memory function, which acts on products in which just one needs to belong to the subnetwork-bulk boundary (see equation (J.3) in the appendix J). In addition, both terms in this product are calculated at the same time of the past, while in the GVA they evolve independently so that the nonlinear memory (as we see from (5.5.13)) keeps track of two different times in the past. Remarkably, we demonstrate that the *combination* of the memory integral and the coloured noise appearing in equation (5.5.12) provides an approximation of the subnetwork reduced dynamics *equivalent* to the one attained by projection techniques, *up to quadratic terms and in the limit of negligible intrinsic noise*. The proof of this equivalence is rather non-trivial, both conceptually and algebraically, thus we refer to section J.5 of appendix J for its entire discussion.

5.6 Application to biochemical networks

In this section we illustrate the GVA reduction method by applying it to a simple toy model and to the EGFR biochemical network from [116], the aim being to assess its accuracy and explaining its computational implementation. We solve in parallel the projected equations from [77], to verify explicitly the equivalence between the two approaches (under the condition specified above of negligible noise) and to compare their precision at $O(\delta \boldsymbol{x}^3)$.

Everywhere in what follows we take $\epsilon \to 0$ and the bulk is chosen initially at steady state, so $\boldsymbol{C}^{\mathrm{bb|s}}(t,t')$, given by (5.5.11), vanishes and $\boldsymbol{\mu}^{\mathrm{b|s}}(t) \equiv \boldsymbol{\nu}(t)$ (the conditional bulk means coincide with the deterministic bulk solutions). Furthermore, the coloured noise of the GVA is identically zero as well as the closed-form expression for the random force in projection methods, which is valid

up to $O(\delta \boldsymbol{x}^2)$ (see section J.3 of appendix J).

5.6.1 Toy model

As a baseline to explain step by step our method, we focus on a simple example, a toy model with 5 species: x_1, x_2 and x_3 belong to the subnetwork, x_4 and x_5 are in the bulk, they undergo two complex association/dissociation reactions. Schematically we write

$$x_1 + x_2 \underset{k^-_{3,12}}{\overset{k^+_{12,3}}{\rightleftarrows}} x_3$$

$$x_1 + x_4 \underset{k^-_{5,14}}{\overset{k^+_{14,5}}{\rightleftarrows}} x_5$$

where $k^+_{12,3}$ and $k^+_{14,5}$ are the association rate constants, while $k^-_{3,12}$ and $k^-_{5,14}$ the dissociation ones. We omit the Langevin noise $\xi_i(t)$ as we consider the $\epsilon \to 0$ limit, which can be achieved by choosing appropriately large reaction volumes, i.e. $V\mu_i \gg 1$ for ordinary states μ_i, $i = 1, ..., 5$. In addition, as we assume the bulk initially at steady state, i.e. $\delta \boldsymbol{x}^{\mathrm{b}}(0) = 0$, we will retrieve a reduced dynamics without random force. To start we shall consider a linearized dynamics around the steady states, thus we look at the deviations w.r.t. the steady states $\delta x_i = x_i - \mu_i$ $i = 1, ..., 5$ which obey a mass action kinetics

$$\begin{aligned}
\frac{d}{dt}\delta x_1 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) \\
&\quad + k^-_{5,14}\delta x_5 - k^+_{14,5}(\mu_1\delta x_4 + \mu_4\delta x_1 + \delta x_1\delta x_4) && \text{(5.6.1a)} \\
\frac{d}{dt}\delta x_2 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) && \text{(5.6.1b)} \\
\frac{d}{dt}\delta x_3 &= -k^-_{3,12}\delta x_3 + k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) && \text{(5.6.1c)} \\
\frac{d}{dt}\delta x_4 &= k^-_{5,14}\delta x_5 - k^+_{14,5}(\mu_1\delta x_4 + \mu_4\delta x_1 + \delta x_1\delta x_4) && \text{(5.6.1d)} \\
\frac{d}{dt}\delta x_5 &= -k^-_{5,14}\delta x_5 + k^+_{14,5}(\mu_1\delta x_4 + \mu_4\delta x_1 + \delta x_1\delta x_4) && \text{(5.6.1e)}
\end{aligned}$$

where we removed the constant terms cancelling at steady states. There are 2 conservation laws, one for the bulk and one for the subnetwork: the total concentration of x_2 and x_3 (similarly for x_4 and x_5) must be conserved, which imply that their fluctuations must be equal and opposite, i.e. $\delta x_2 = -\delta x_3$ and $\delta x_4 = -\delta x_5$.

Reduced dynamics

There is one boundary species, δx_1 (which interacts with the bulk species δx_4 and δx_5): its dynamics is thus the only one affected by memory effects. By applying the formulas for the memory

(5.4.6) and (5.5.13) we obtain the effective equation for $\delta x_1(t)$: let us show it step by step. While for the linear memory we can simply apply formula (5.4.6), for the nonlinear corrections we shall start from the cubic terms in the action that contain at least a bulk species, to be expressed conditionally on the subnetwork boundary species (i.e. just δx_1)

$$\Delta\mathcal{H} = \int_0^T dt\Big[k^+_{14,5}\mathrm{i}\hat{x}_1(t)\delta x_1(t)\delta\mu_{4|1}(t) + k^+_{14,5}(\mathrm{i}\hat{\mu}_{4|1}(t)\delta\mu_{4|1}(t)\delta x_1(t) - \mathrm{i}\hat{\mu}_{5|1}(t)\delta\mu_{4|1}(t)\delta x_1(t))\Big] \quad (5.6.2)$$

We need the explicit expression for the conditional means $\delta\mu_{4|1}(t)$, $\hat{\mu}_{4|1}(t)$ and $\hat{\mu}_{5|1}(t)$, which can be found by applying (5.5.9) and (5.5.7)

$$\delta\mu_{4|1}(t) = -\int_0^t dt' e^{-(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')}k^+_{14,5}\mu_4\,\delta x_1(t') \quad (5.6.3)$$

$$\mathrm{i}\hat{\mu}_{4|1}(t) = -\int_t^T dt' e^{(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')}k^+_{14,5}\mu_1\,\mathrm{i}\hat{x}_1(t') \quad (5.6.4)$$

$$\mathrm{i}\hat{\mu}_{5|1}(t) = \int_t^T dt' e^{(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')}k^-_{5,14}\,\mathrm{i}\hat{x}_1(t') \quad (5.6.5)$$

Once substituted in (5.6.2), we can read from the effective action an equation for $\delta x_1(t)$ - formally what multiplies $\mathrm{i}\hat{x}_1(t)$

$$\begin{aligned}\frac{d}{dt}\delta x_1 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) + \int_0^t dt' M_{11}(t-t')\delta x_1(t') \\ &+ \int_0^t dt'\int_{t'}^t dt'' M_{11,1}(t,t'',t')\delta x_1(t')\delta x_1(t'')\end{aligned} \quad (5.6.6)$$

with

$$M_{11}(t-t') = (k^-_{5,14} + k^+_{14,5}\mu_1)k^+_{14,5}\mu_4\, e^{-(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')} \quad (5.6.7)$$

and

$$M_{11,1}(t,t'',t') = k^{+2}_{14,5}\mu_4\, e^{-(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')}(\delta(t-t'') - (k^-_{5,14} + k^+_{14,5}\mu_1)) \quad (5.6.8)$$

corresponding to (5.4.6) and (5.5.13). As we have explained in section 5.5.2, solving the integro-differential equation (5.6.6) via only differential equations requires $2N^{\mathrm{b}}$ additional variables. The number of bulk species is 2 (x_4 and x_5), but they can be reduced to 1 because of the conservation law, thus in total 2 additional variables must be introduced, one of them simply giving the conditional mean (5.6.3) and satisfying

$$\frac{d}{dt}\delta\mu_{4|1} = -k^+_{14,5}\mu_4\delta x_1 - (k^-_{5,14} + k^+_{14,5}\mu_1)\delta\mu_{4|1} \qquad \delta\mu_{4|1}(0) = 0 \quad (5.6.9)$$

The other one, z, simply following the procedure outlined in appendix K, is the solution of

$$\frac{d}{dt}z = -k^+_{14,5}\delta x_1\delta\mu_{4|1} - (k^-_{5,14} + k^+_{14,5}\mu_1)\,z \qquad z(0) = 0 \quad (5.6.10)$$

where $-(k^-_{5,14} + k^+_{14,5}\mu_1)$ is the only nonzero eigenvalue of $\boldsymbol{K}^{\mathrm{bb}}$. These additional equations must be solved jointly with (5.6.1b), (5.6.1c) (the internal subnetwork equations in their original form) and

$$\frac{d}{dt}\delta x_1 = k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) - (k^-_{5,14} + k^+_{14,5}\mu_1)\delta\mu_{4|1} - k^+_{14,5}\delta x_1\delta\mu_{4|1} + (k^-_{5,14} + k^+_{14,5}\mu_1)\,z \tag{5.6.11}$$

We solved numerically (5.6.9), (5.6.10) and (5.6.11) to find the time evolution of $\delta x_1(t)$; in figure 5.1 (left) we plot a related dimensionless quantity, the fractional concentration deviation, defined as $\delta\tilde{x}_1 = (x_1(t) - \mu_1)/\mu_1$. To remark the increased accuracy achieved by including memory terms, we contrast time courses obtained in our reduced model to the ones from other usual and simpler approximations for biochemical networks: one in which the subnetwork is considered as isolated and one in which the bulk is assumed to be fast, thus able to reach the steady state on the timescale of subnetwork's evolution (steady state bulk). Time courses given by the nonlinear GVA are visually indistinguishable from the exact ones, as opposed to the other approximation whose performance is less accurate. Invoking these simplifications is typically justified by timescales separation or by a weak coupling between subnetwork and bulk, i.e. by conditions sufficient to assume their evolution approximately independent. If their dynamical coupling is made much weaker (e.g. by scaling down $k^+_{14,5}$ and $k^-_{5,14}$ by a factor 100) one would obtain that all the approximate schemes perform excellently, with time courses completely overlapping. In the case of a stronger coupling, the two alternative approximations, in particular the steady state bulk, lead to poorer predictions - see inset of figure 5.1 (left). A systematic model reduction, as provided by our method and by projected equations, improves the agreement with the full solution *regardless* of any assumption on the strength of bulk/subnetwork interactions, thus it ensures a greater flexibility of application. In partially observed biological systems the criterion to select a certain subsystem may actually depend on particular cases of analysis and on the available data.

Quantitative tests

We next look at the error of approximation. In analogy with [77], we make the initial conditions scale by a factor which measures the initial deviation w.r.t. to the steady state; to estimate its overall magnitude, we take the initial root mean squared deviation $\delta = (\{\sum_s [\delta\hat{x}_s(0)]^2\}/N^{\mathrm{s}})^{1/2}$. We quantify the accuracy of the approximation by

$$\Delta = \frac{1}{T}\int_0^t dt' \frac{1}{N^{\mathrm{s}}}\sum_{s=1}^{N^{\mathrm{s}}} |\delta\tilde{x}_s(t) - \delta\hat{x}_s(t)| \tag{5.6.12}$$

$\delta\hat{x}_s(t)$ being the exact solution for the subnetwork. Δ is an absolute deviation in the dimensionless concentration of each subnetwork species, averaged over species and a time course of $T = 15$ s, chosen to entail the transient regime. From figure 5.1 (right) one sees that the error for the simpler (Markovian) approximations is substantially larger in absolute terms, a conclusive evidence that non-Markovian terms as memories ensure improved predictions. In addition, their error is linear, while it scales cubically for nonlinear projection methods and GVA. They both capture quadratic observables consistently but account for terms $O(\delta^3)$ just partially and not systematically: the different way in which these are included explains the different accuracy of the two approaches. In the nonlinear projection methods, the error is due to the fact that there are cubic contributions left from the random force which do not vanish if the bulk is initially at steady state. In GVA, the error is due to the fact that we truncate the perturbative expansion after $O(\Delta\mathcal{H})$, while to be systematic at the cubic order in the dynamics we would need also fourth order terms in the effective action (i.e. from $O(\Delta\mathcal{H}^2)$). We also plot the error of an approximation (denoted as "$O(\delta^2)$ only") in which products evolve solely under the linearized dynamics, i.e. without memory terms (which contribute to the δ^3 order). We see that the accuracy is lower, as expected, but the error still scales cubically as this case corresponds to a systematic elimination of all the $O(\delta^3)$ terms appearing in the dynamics of quadratic observables.

5.6.2 Application to EGFR

We consider the network of protein-protein interactions around EGFR as in the model by Kholodenko et al. [116], already analyzed as a relevant testbed for the projection methods in [77]. The reaction network is shown in figure 5.2 and we refer to [77] for a full list of species and their abbreviations. The dynamics is a result of the law of mass action, with kinetic parameters from [116], which can be written for the deviations from the steady states $\delta\boldsymbol{x}(t)$. For the sake of the comparison with [77], we also choose the same initial conditions (which maximize nonlinear effects consistently with the conservation laws in the subnetwork). One needs to add 3 enzymes to account for the Michaelis-Menten type of reactions in the network (species ER, EP, ES in figure 5.2). The bulk is defined by Src homology and collagen domain protein (Shc) and any complexes containing it. They interact with 4 subnetwork species, which therefore form the "boundary", i.e. the species exhibiting memory effects: phosporylated EGFR (denoted as RP), growth factor receptor-binding protein 2 (Grb2), Son of Sevenless homolog protein (SOS) and protein-complex Grb2-SOS (GS). All the other "interior" subnetwork species obey the initial mass action dynamics.

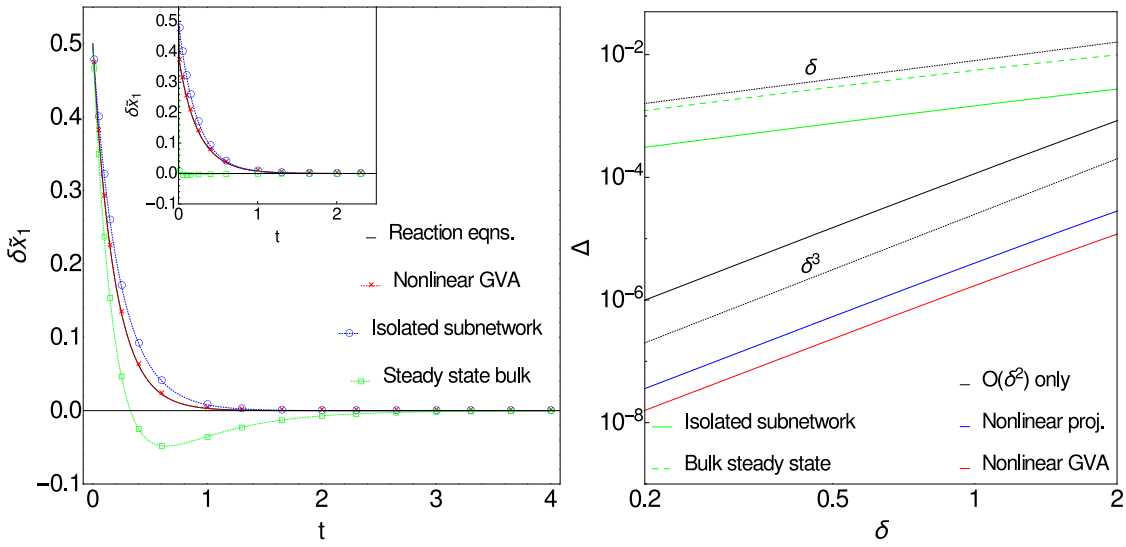

Figure 5.1: (Left) Time courses of $\delta\tilde{x}_1$ given by the nonlinear GVA, the isolated subnetwork and steady state bulk approximations compared to the exact ones. (Right) Error of approximation, Δ, as a function of the overall initial deviation from steady state δ; the nonlinear GVA is compared to projection methods (for both the error scales as δ^3) and the two simpler approximations (whose error grows already as δ). Rates and steady states are set to $k^+_{12,3} = k^+_{14,5} = 1$, $k^-_{3,12} = k^-_{5,14} = 2$ and $\mu_1 = \mu_2 = \mu_4 = 1$, $\mu_3 = \mu_5 = 1/2$. In the inset on the left, the strength of subnetwork-bulk interaction is increased by setting $k^+_{14,5} = 5$, $k^-_{5,14} = 10$. Initial conditions are chosen as $\delta\tilde{x}_1(0) = \delta\tilde{x}_2(0) = 1/2$, $\delta\tilde{x}_3(0) = -1$, $\delta\tilde{x}_4(0) = \delta\tilde{x}_5(0) = 0$ (the bulk is initially at steady state) for the left figure, while on the right we vary the factor tuning the initial deviation from steady states.

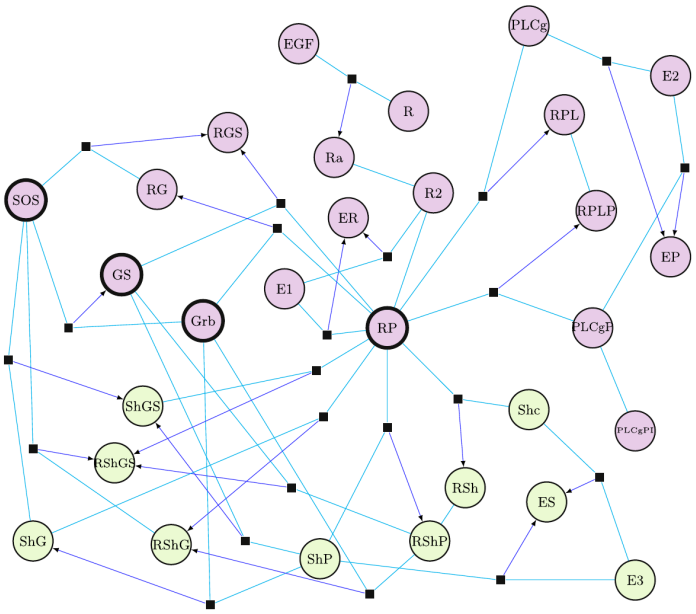

Figure 5.2: Figure from [77] depicting the reaction network from [116]. The subnetwork is highlighted in light red while bulk species appear in green.

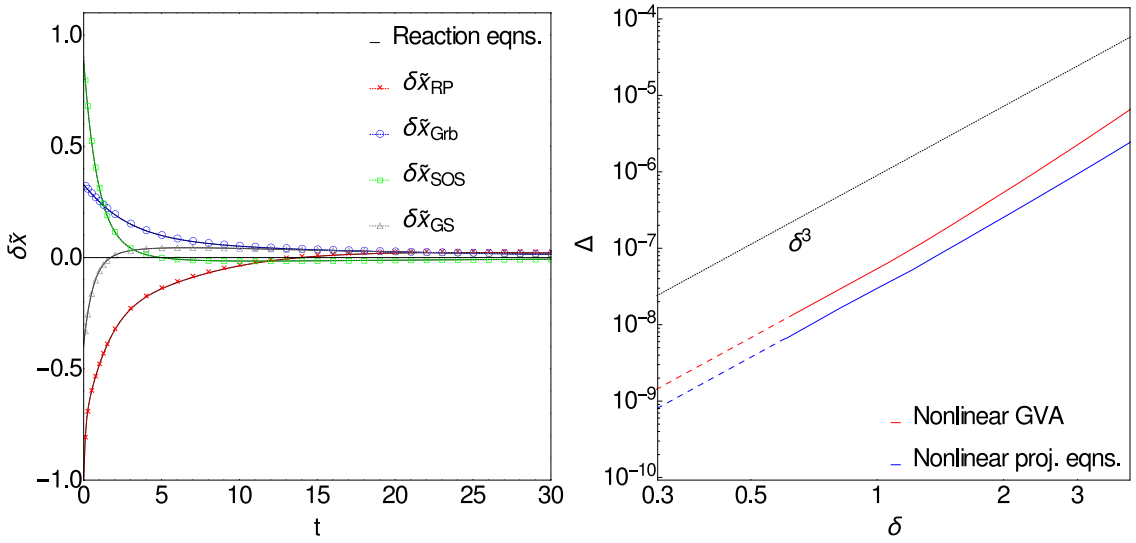

Figure 5.3: (Left) Time courses of the fractional concentration deviations from steady state for the 4 subnetwork boundary species RP, Grb, SOS, GS. (Right) Error of approximation, Δ, as a function of the overall initial deviation from steady state δ; the nonlinear GVA is compared to nonlinear projected equations (for both the error scales as δ^3). The dashed lines indicate an interpolation of the error where it cannot be numerically estimated with such precision. Rates, steady states and initial conditions are set as in [77].

Quantitative tests

We apply both the projection methods as in [77] and the nonlinear GVA reduction method here developed for the sake of comparison. In particular, to focus on the contribution of memories, the initial deviations from the steady state in the bulk are zero and the intrinsic noise is neglected, as for the toy model. As we show in figure 5.3 (left) for fractional concentration deviations $\delta\tilde{x}_i = (x_i - y_i)/y_i$, a very accurate prediction of subnetwork time courses can be achieved by considering memory terms. Under the condition of vanishing noise, the performance of the nonlinear GVA is equivalent to the one of projection methods to the quadratic order (section J.5 of appendix J), thus for a thorough, quantitative investigation of its improved accuracy compared to other approximations (steady state bulk and isolated subnetwork) see [77].

An open question is nevertheless to what accuracy the two methods capture the *cubic* terms: this can be assessed by developing a quantitative test as in section 5.6.1, with the initial root mean squared deviation Δ defined by (5.6.12) (here we consider $T = 150$ s to capture the transient regime). The value of Δ as a function of the initial deviation from steady state δ is visible in figure 5.3 (right), for the nonlinear GVA and the nonlinear projection methods. As expected, Δ grows as δ^3 for both the methods, the difference (of roughly a factor 2) is explained by the

fact that they treat cubic terms differently and *not systematically*. As a consequence, the overall accuracy depends on the particular network considered: for the EGFR one, projected equations are more accurate whereas for the toy model we had the opposite situation (see figure 5.1 (right)). Let us remark that in absolute terms the error for the EGFR system is extremely small for both model reduction strategies, making their respective time courses visually indistinguishable. One could thus consider to prioritize the computational advantage offered by the nonlinear GVA over the higher accuracy of projection methods given that biochemical concentrations might not be characterized experimentally to such degree of accuracy. In terms of numerical implementation, in fact, one needs 20 additional variables for the nonlinear GVA as opposed to the 255 required by projection methods (see appendix K and [77]).

Nonlinear memory function

We next analyze the nonlinear memory function $\boldsymbol{M}^{\mathrm{s,ss}}(t,t'',t') \equiv \boldsymbol{M}^{\mathrm{s,ss}}(t-t'',t-t')$, which depends on two time differences, $t-t'$ and $t-t''$, and whose expression can be extracted from (5.5.13) for $t'' > t'$. We symmetrize this expression w.r.t. the role of t' and t'' to avoid imposing a defined time ordering between them (nevertheless $t', t'' < t$).

The first term in (5.5.13) contains a $\delta(t-t'')$, thus it gives contributions only for $t = t''$, what we could call the "edge" in 3-dimensional representation of this function. The second and third lines describe the memory for all the values of $t' \neq t''$, so the "interior"' part of such a representation. We denote these as GVA-edge and GVA-interior terms; we plot the edge term simply as a function of $t-t'$ by dropping the δ piece - see figures 5.4 and 5.5 (left) - while examples of the interior terms are shown in the 3D figures 5.4 and 5.5 (right). In particular, we consider a version of the nonlinear memory function in (5.5.13) made dimensionless w.r.t. concentrations, as it would appear in the equations for the fractional concentrations $\delta\tilde{\boldsymbol{x}}$s (e.g. the element $\tilde{M}_{\mathrm{SOS,SOS\text{-}SOS}}(t-t',t-t'') = y^2_{\mathrm{SOS}} M_{\mathrm{SOS,SOS\text{-}SOS}}(t-t',t-t'') y^{-1}_{\mathrm{SOS}}$).

We compare the GVA nonlinear memory function with the one of the projection approach (see (J.3.5) in appendix J), which is a function of just one time difference $t-t'$. In figures 5.4 and 5.5 (left) we plot two examples, the self-memory and a cross-memory function for SOS, against the GVA memory terms for the same species (where we set $t' = t''$ in the interior piece). As the bulk is initially at steady state and the intrinsic noise is neglected, the overall GVA nonlinear memory on a certain species is equivalent, up to $O(\delta\boldsymbol{x}^2)$, with the projection one. They both can be seen as sums of contributions given by the memory functions applied to the set of products. Importantly, the projection memory function applies to *all* subnetwork products containing at

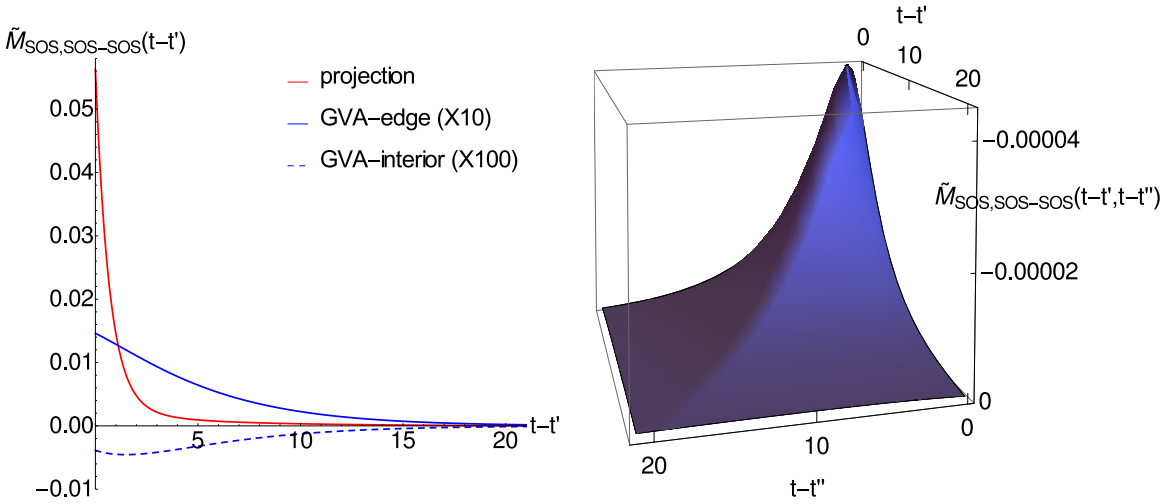

Figure 5.4: Nonlinear self-memory of subnetwork species SOS. (Left) Comparison of GVA-edge and GVA-interior terms with projection methods; the GVA-interior term is taken along the diagonal, i.e. for $t' = t''$. The GVA-edge and interior terms are multiplied respectively by a factor 10 and 100 to make them visible on the scale of the projection memory, whose amplitude is significantly bigger. Note nevertheless that in principle the interior term would not be directly comparable with the nonlinear memory from projection methods and the GVA-edge term since they do not have the same units of measurement. The total nonlinear memory $\tilde{\mathcal{M}}_{SOS}(t)$ has dimensions t^{-1}, thus the nonlinear projection memory function, as it is integrated once over time, has units t^{-2}. The GVA-edge has the same dimensions (it is integrated over time twice but the delta contributes with t^{-1}); on the other hand, the interior term, appearing in a double time integral, has units t^{-3}. (Right) The GVA-interior memory function is plotted in 3D to show its shape with $t' \neq t''$: we see that the diagonal contribution (i.e. for $t' = t''$) is relatively dominant.

least one boundary species while the GVA one is defined only for products of *boundary* species and this helps explain the significant difference in amplitude visible in figures 5.4 and 5.5 (left). In this sum, the projection "distributes" the memory more into positive and negative terms from all the different products of subnetwork species, while in the GVA all the interior subnetwork species are effectively expressed as functions of the boundary ones, resulting in fewer terms.

Finally, in figure 5.6, we show the entire nonlinear memory on SOS, plotting separately the pieces from the integral of the GVA-edge term and of the GVA-interior one respectively: they both contribute significantly depending on time t.

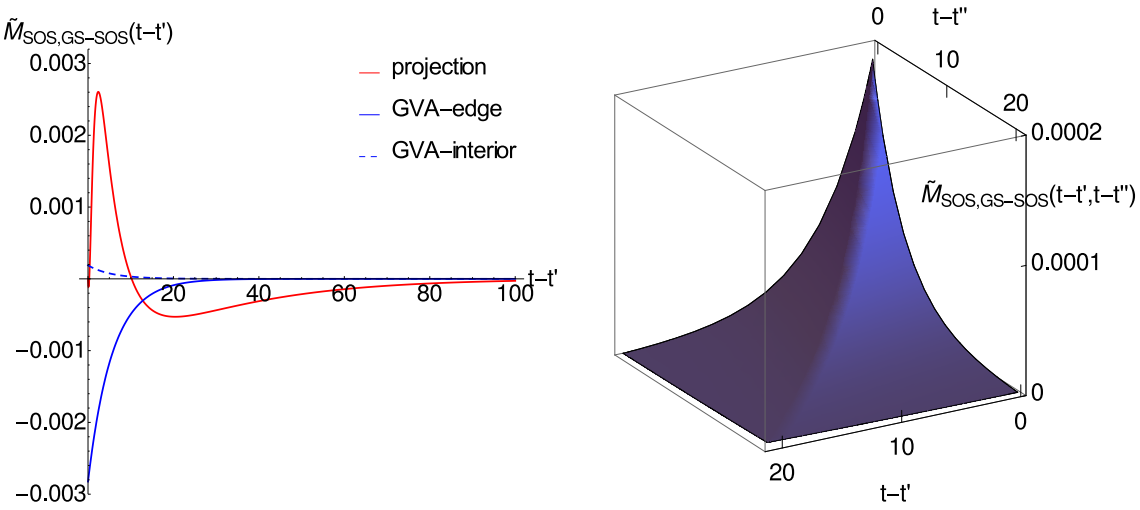

Figure 5.5: Example of nonlinear cross-memory acting on SOS, defined for the product GS-SOS. (Left) Comparison of GVA-edge and GVA-interior (for $t' = t''$) terms with projection methods memory. Similarly to $\tilde{M}_{\text{SOS,SOS-SOS}}(t-t')$ in figure 5.4 (left), the GVA nonlinear memory function $\tilde{M}_{\text{SOS,GS-SOS}}(t-t')$ is smaller in amplitude w.r.t. its projection analogue because effectively it accounts for what in the projection approach is a sum of terms with positive or negative sign - the nonlinear memories on SOS, defined for all subnetwork products containing at least one boundary species but where one could imagine expressing the interior species in terms of GS and SOS (both boundary ones). (Right) Full GVA-interior term as a function of $t-t'$ and $t-t''$: as in figure 5.4 (right), it is somehow "peaked" around $t' = t''$ (but not so much to say that it is concentrated only along the ridge).

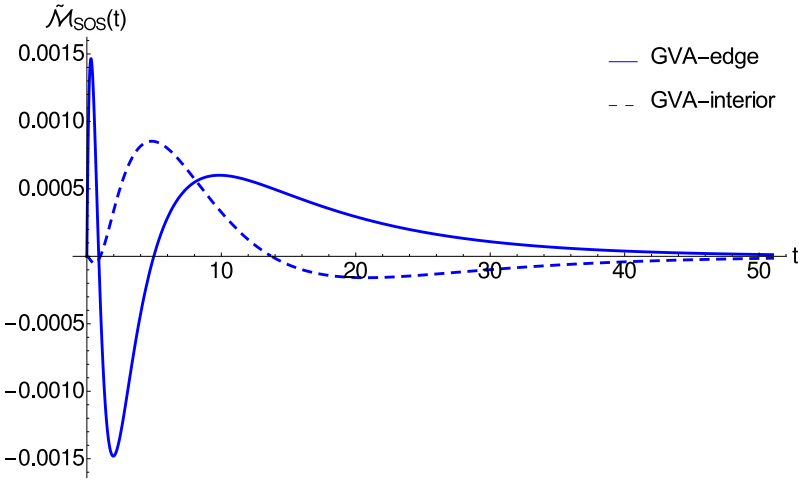

Figure 5.6: The integrals of the GVA-edge and GVA-interior terms as they appear in the entire nonlinear memory of SOS. The size of their contribution is comparable and changes notably with time.

5.7 Discussion and conclusion

In this chapter we have developed a model reduction technique based on a Gaussian approximation of the CLE describing complex formation and dissociation in biochemical networks. We have first re-derived a Gaussian approximation variationally via the MSRJD formalism, which allows a generalization to the case in which the diffusion term in the true and the approximating process are not the same. It is known for real distributions [110] that with a mismatch of those terms the KL diverges, while here, looking at just the stationary points of KL between complex distributions, we obtain that the variational noise covariance is the average of the true one w.r.t. the approximating Gaussian distribution (see (5.3.11)): it would be interesting to see whether such a result could be derived more rigorously from just real distributions.

Essentially the GVA yields an effective linearization around time dependent means. Their evolution (see (5.3.8)) does not fully correspond to macroscopic trajectories as it depends on correlations as well, thus it contains corrections stemming from fluctuations in systems of finite size. Therefore it is more accurate than the LNA and they become equivalent only in the infinite volume limit. In other words, the Gaussian approximation acts as a 2 moment-closure approximation, whose error on the means can be shown to scale as $1/V^2$, thus it is comparable to the error of the CLE and smaller than the LNA one (which scales as $1/V$, see [127, 145]).

For the purpose of model reduction, one basically needs the full correlation matrix (comprising bulk-subnetwork blocks) to retain all the relevant dynamical information from the bulk. In addition, the GVA is usable primarily with constant means, at least in the bulk, to avoid having to solve multiple time integrals; this scenario seems also well justified according to how we posed the problem, i.e. we would not expect to have strong prior knowledge about the bulk which characterizes it far from the steady state.

The subnetwork reduced description is known to have the structure of a generalized Langevin equation, i.e. with additional non-Markovian terms (a memory and coloured noise). We started from the linearized dynamics given by the GVA, which allowed us to find the reduced dynamics by direct elimination. In signalling pathways, linear models describe the so called *weakly activated* cascades and are considered of theoretical interest for building coarse-grained descriptions [146, 147]. We next developed a perturbative expansion of the effective action to estimate the nonlinear corrections and for the whole discussion, we kept the comparison with projection techniques as a baseline to verify the consistency and the accuracy of our method. For example, we noted that implementing the subnetwork- bulk split in the GVA is more easiliy generalizable to time

dependencies (i.e. to the linearization around time dependent means).

In the limit $\epsilon \to 0$, we demonstrated that the reduced GVA reaches the same level of accuracy up to $O(\delta \boldsymbol{x}^2)$, yet one does not need to project the dynamics onto an enlarged space containing also products, which yields a proliferation of variables when large networks are considered. We then presented an easily understandable illustration of our method with a toy model, which can be fully treated analytically. We provided quantitative evidence of the better performance achieved by accounting explicitly for memory effects compared to other usual model reduction schemes (isolated subnetwork and steady state bulk). These are known not to capture appropriately the time courses when subnetwork and bulk evolve on similar timescales. Timescales separation allows one to *postulate* the validity of simplified descriptions, while our procedure is conceived to *derive* systematically reduced models regardless of the relative speed bulk-subnetwork (no assumption in this regard is made throughout our derivation). The choice of a subnetwork can be therefore flexibly adjusted to particular cases, depending on the available information or the particular feature to analyze. The main application for our model reduction strategy are protein-protein interaction networks: we showed for such a model (the EGFR network from [116]) that the agreement with the exact curves is excellent, the absolute error being slightly higher than for projection methods but still extremely small. Such improvements in the accuracy of prediction are an essential contribution to build realistic models of biochemical networks and interpret correctly data.

Interesting connections can be established to other approaches [137, 138] where the *marginalization* w.r.t. the environment enables to uncouple the dynamics of a subsystem but still to embed environmental effects. The starting point is a more general framework, describing reactions as continuous time Markov Chain obeying the Chemical Master Equation. The environment is modelled as unidimensional stochastic variable which exerts a feed-forward influence, i.e. the environment modulates dynamically the subsystem but not viceversa. The marginalization replaces the original dependence on the environment by its mean conditional on the full history of the subsystem. Consequently, the resulting marginal dynamics of the subsystem is non-Markovian, i.e. it satisfies a generalized master equation with memory, and *self-exciting*, i.e. the past values feed back onto the present ones. The conditional expectation of environment can be found as the solution to a stochastic *filtering* problem where one can imagine reconstructing a hidden stochastic process (the environment) at time t from from the "observed" (fixed) values of the subnetwork. A practical limitation of these approaches is that the implementation depends on finding suitable approximate schemes for the marginal and conditional dynamics, as most often they cannot be computed analytically. The key ingredient, the marginalization w.r.t. the environment by conditioning on the

subnetwork history, is in common with what we presented here; yet we assume a specific form (a Gaussian) for the path probability distribution. In this way the conditional means and covariances are the ones that one could obtain by Kalman filter [63], i.e. the paradigmatic filtering technique for linear problems. The advantage is to obtain analytical expressions for the effective non-Markovian terms (memory and coloured noise), which emphasize the feed-forward and feed-back exchange between subnetwork and bulk.

Model reduction is a tool for partitioning consistently and systematically the network into modules that are easier to study (in our case without invoking a timescales separation). A feature worth remarking of reduced descriptions is then their *scalability*: when the environmental part is no longer to be simulated, the computational effort can be significantly decreased. Apart from the computational convenience of model reduction, we stress also the *conceptual* importance of studying how a cellular subsystem couples to the rest of the cellular environment. It has been for example recently shown that equilibrium-like domains might arise in non-equilibrium systems [148] by considering generic bipartite graphs, a scenario akin to our subnetwork/bulk split. Analogously, techniques of systematic model reduction can explain the emergence of dissipative behaviours out of microscopic conservative dynamics [149]. Our framework thus lends it-self to a controlled investigation of how the environment might affect subparts of the system via perturbations or fluxes which drive them out of equilibrium, with the possibility, ultimately, of controlling them experimentally.

Furthermore, our formula for the covariance of the coloured noise (5.4.7) shows explicitly the decomposition of stochasticity into different contributions, i.e. intrinsic noise, inherent to the random timing of biochemical events, and the extrinsic noise, variation due to the interaction with the environment [8, 150, 151]. As clear from our way of proceeding, conditioning was a crucial step, which allowed us to reduce the description to a subset of trajectories and to exploit the available information. The importance of conditional expectations to distinguish sources of noise has already been highlighted in several recent studies [138, 150, 152, 153], e.g. to understand how noise in gene expression may depend on processes extrinsic to that, as turnover ribosomes or cellular growth. Typically, the law of total variance is employed to decompose the fluctuations on some biochemical species into contributions intrinsic to it and contributions coming from either a static [150] or dynamic [138, 152] environment - the latter requiring to condition on entire environmental histories. A systematic generalization [153] to any group of variables, achieved by successive conditioning on their full histories, was translated then into precise conditions for reporter experiment design. Here we derived a novel way to track the dynamic propagation of noise which could be then relevant to identify and ultimately measure the components of cellular heterogeneity.

Several aspects in this regard deserve further, separate investigation. A very important and promising feature of this approximation is that it provides not the conditionally averaged evolution but the reduced evolution fully accounting for stochastic effects. More precisely, the GVA reduced equations retain the original subnetwork intrinsic randomness, its propagation across the bulk and the one stemming from the "ignorance" about the initial conditions in the bulk. Starting from section 5.5.3, we investigated the limit of small noise for the sake of a comparison with the projection methods and to place the emphasis on stochastic terms arising from uncertain initial conditions. We think that this restriction could be easily removed at the price of longer expressions. This would yield a mathematically rigorous, analytically controlled tracking of stochastic effects, which would be decomposed into their fundamental components in a *principled* way. Reduced models including stochastic terms pose additional computational challenges but are definitely worth investigating, as it is fully acknowledged that biochemical reactions are better described by stochastic equations and fluctuations themselves are likely to play a crucial role towards biological functions (see [154] and references therein).

A separate forthcoming paper will address several related questions. For example, equation (5.4.7) points out that the extrinsic noise statistics, including temporal correlations, can be derived from the statistics of the initial states of the bulk and the one of white noise: importantly, this makes possible to test assumptions about extrinsic noise. In this regard, the GVA would be a more powerful tool in comparison to projection methods, since one could analyze different variants not only in terms of choices of initial conditions but also of white noise statistics.

Let us sketch more precisely some ideas for this future project, which could follow the baseline provided in [134], thus including as well a comparison with projection techniques. One would need to make assumptions about the distribution of the initial values of the bulk and the subnetwork: the simplest case would be independent Gaussian fluctuations, with an overall scale set by a parameter ϵ_0 (independent from the ϵ governing the volume size, see [134]). This choice would allow one to consider cases where there is substantial uncertainty about the initial state but the copy number noise in the dynamics is still low. For the general comparison between the nonlinear noises in the GVA and in the projection approach, one could think to what accuracy, in orders of ϵ_0, we would expect agreement. It is likely that within the projection method we cannot determine e.g. the $O(\epsilon_0^2)$ term in the random force correlator consistently, while we expect it to be possible in the GVA, as it expands directly the coloured noise correlator. In fact, the projection approach can so far predict the linear and quadratic (in the initial values $\delta \boldsymbol{x}^{\mathrm{b}}(0)$) contributions to the random force. A product of random forces, as needed for a correlator, after averaging,

would be $O(\delta \boldsymbol{x}^{\mathrm{b}}(0)\delta \boldsymbol{x}^{\mathrm{b}}(0))$ and $O(\delta \boldsymbol{x}^{\mathrm{b}}(0)^2\delta \boldsymbol{x}^{\mathrm{b}}(0)^2)$, giving respectively $O(\epsilon_0)$ and $O(\epsilon_0^2)$ contributions. However, if we accounted for $O(\delta \boldsymbol{x}^{\mathrm{b}}(0)^3)$ terms in the random force, then these would give $O(\delta \boldsymbol{x}^{\mathrm{b}}(0)\delta \boldsymbol{x}^{\mathrm{b}}(0)^3)$ contributions to the random force correlator that are as well $O(\epsilon_0^2)$. Therefore, it seems plausible that some terms are not exhaustively and systematically computed in the projection approach; given that, one could rigorously verify first that the GVA accuracy really is higher if we truncate the expansion at a certain order and then assess what impact this has on the accuracy of the description of the subnetwork dynamics.

Furthermore, we note that the spectrum of the bulk determines the modes of the memory function, thus it can help identify different timescales: this aspect can be exploited for additional simplifications. Timescales separation typically leads to the automatic identification of modules operating at different speed in some biological networks (as the ones involving gene regulation), each amenable to be chosen as a "subnetwork". Timescales variation could also suggest and guide other ways of coarse-graining the description. For example, estimating in advance the regime of time differences where the memory kernel is significant, one could think of short memory approximations, as in [143], where the memory is explicitly taken into account only for the species strongly driven by non-Markovian effects. Or also, renormalizing reactions by typical decay times of the memory would allow one to map integro-differential equations into an effective Markov description [155].

Apart from the information on typical timescales, it is likely that spectra of the network weights (couplings) matrices crucially determine several features of the reduced dynamics. For example, complex eigenvalues give rise to oscillations, phenomena of strong localization of the non-Markovian effects can stem as a result of inhomogeneities in the connectivity (from the memory structure (5.4.4) one sees that the way in which the memory picks up the eigenvalues of $\boldsymbol{K}^{\mathrm{bb}}$ depends on *local* interactions $\boldsymbol{K}^{\mathrm{bs}}$). Graph spectral properties actually give evidence of modularity, localization, existence of slow modes and other interesting structural and dynamical features [156–158] all of which can suggest model-specific ways of analysis.

Another plausible scenario is the one in which the knowledge on the bulk is even more restricted, i.e. we do not know precisely the connectivity structure but we can estimate via some prior information statistical properties as the degree of symmetry or the sparsity: sampling from this distribution would allow one to explore how the subnetwork-bulk connectivity governs the subnetwork reduced dynamics.

Finally, the novelty of the framework we presented opens many directions for future developments, on the front of dynamical modelling and inference; one starting point could be lifting some of the assumptions we made to better fit the features of specific networks.

For the sake of full analytical treatment, we assumed an approximating Gaussian distribution (i.e. peaked around its average value and with small fluctuations around it) and it may not be adequate in some contexts: for example, biochemical networks might exhibit bi- or multistable states or oscillations. A Gaussian approximation is expected to provide a good approximation with a unique steady state, while it does not capture multimodality and also large deviations. A first approximation to capture multistability could be given by Gaussian mixture models [159], but it could be worth also trying to overcome the restriction to the Gaussian scenario. In this regard, the general variational formalism put forward by Eyink [118], could serve as a reference framework.

Furthermore, there exist other types of nonlinearities which are not captured by the set of binary reactions we considered here, such as enzymatic dynamics and Hill functions. Although not straightforward, we expect that appropriate changes can lead us to extend the applicability of the GVA to those settings. One could follow a procedure similar to the one used to treat Michaelis-Menten kinetics by projection methods [135]: linearize the equations to get a mass action form, apply the reduction technique, and then reinstate the full nonlinear dynamics.

Let us now comment on the assumptions for fluctuations. We started from a continuous approximation of the ME, the CLE, whose fundamental premise of validity is a large population size ($\sim$ hundreds to thousands of molecules). On the other hand, it is widely established that fundamental biomolecules such as DNA and mRNA, are present in small numbers [8, 9], so that fluctuations, which scale as the inverse square root of this number, can gain a significant size, even of order unity. As a first step to improve the accuracy of description when fewer molecules are considered, one could include higher order finite volume corrections (for example by the analytical method based on Feynman diagrams developed in [160]). We then know that small populations (tens to hundreds) are more realistically described by ME (discrete variables); in this case, one has to appeal to other dynamical functional approaches, such as the well known Doi-Peliti formalism [161–163], which entails a fundamental connection of stochastic processes (as stochastic gene expression [119]) to many body problems. This is an additional hint that analogies with theoretical physics and quantum mechanics can be mathematically powerful and conceptually insightful in the biological arena and, in our opinion, should be increasingly exploited.

The need to incorporate delays into models has been acknowledged also in assembly processes (translation and transcription) [12]: a Markov chain model assumes that reactions occur instantaneously once the molecules reach the right configuration, which is proven to be wrong in this case. This evidence suggests another idea for future work, the inclusion of the gene expression level:

here it was considered slow enough to be assumed at steady state on the time scale of protein-protein interactions, which are faster and relatively homogeneous. Significant timescales variation in gene expression, due to different lifetimes of species involved (such as mRNAs and proteins), should be carefully taken into account in the formalism. For instance, it could enter perturbative parameters in non trivial way, justifying the construction of perturbative approximation techniques especially tailored to reactions with two different (fast/slow) timescales [164].

Finally, as a natural continuation of this analysis, one could think of the subnetwork being observed and explore the inverse direction: once an insight into reduced subnetwork dynamics has been established, what can we infer about the bulk? Provided for example that typical properties of memory functions can be measured (at least indirectly), their analysis would enable the inference of bulk configuration and/or dynamical behaviour. In this perspective, fixing and conditioning on the subnetwork provides the way to incorporate observations in our framework and derive posterior distributions, similarly to what has been done in chapters 3 and 4. They could thus be used to develop efficient techniques of Bayesan inference of the dynamics from data. Other approximate approaches were conceived for this task (based on variational methods [80, 110, 121, 165] or system size expansions [99]) and turned out to be conceptually advantageous with respect to MCMC sampling inference methods for intracellular stochastic kinetics.

Appendix G

Gaussian Variational Approximation

G.1 Gaussian Variational Approximation

We denote $\boldsymbol{y} = (\boldsymbol{x}, -\mathrm{i}\hat{\boldsymbol{x}})$ in such a way that the joint "distribution" of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ can be written

$$P(\boldsymbol{y}) = \frac{e^{\mathcal{H}(\boldsymbol{y})}}{Z} P_0(\boldsymbol{x}) \tag{G.1.1}$$

$P_0(\boldsymbol{x})$ being the unknown initial probability distribution. We then proceed with the variational approximation by assuming that the approximating distribution is Gaussian

$$P(\boldsymbol{y}) = \frac{e^{\mathcal{H}(\boldsymbol{y})}}{Z} P_0(\boldsymbol{x}) \approx \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_{\text{gen}}, \boldsymbol{C}_{\text{gen}}) = Q(\boldsymbol{y}) \tag{G.1.2}$$

The Gaussian $\mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_{\text{gen}}, \boldsymbol{C}_{\text{gen}})$ is completely determined by two sets of parameters, the vector of mean values $\boldsymbol{\mu}_{\text{gen}} = \langle \boldsymbol{y} \rangle_Q$ and the covariance matrix $\boldsymbol{C}_{\text{gen}} = \langle \boldsymbol{y}\boldsymbol{y}^T \rangle_Q$. We define the Kullback-Leibler (KL) Divergence [114] between P and Q

$$\begin{aligned}
\mathrm{KL}(Q\|P) &= \int D\boldsymbol{y}\, Q(\boldsymbol{y}) \ln \frac{Q(\boldsymbol{y})}{P(\boldsymbol{y})} = \\
&= \left\langle -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_{\text{gen}})^T \boldsymbol{C}_{\text{gen}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\text{gen}}) - \frac{1}{2} \ln (2\pi)^d \det(\boldsymbol{C}_{\text{gen}}) + \mathcal{H}(\boldsymbol{y}) + \ln Z - \ln P_0(\boldsymbol{x}) \right\rangle_Q = \\
&= -\frac{d}{2} - \frac{1}{2} \ln \left[(2\pi)^d \det(\boldsymbol{C}_{\text{gen}}) \right] + \ln Z - \langle \mathcal{H}(\boldsymbol{y}) + \ln P_0(\boldsymbol{x}) \rangle_Q
\end{aligned} \tag{G.1.3}$$

where d is the dimension of the vectors involved, i.e. $d = N(2T/\Delta + 1)$ and Δ is the elementary time step. With real measures, $\mathrm{KL}(Q\|P) \geq 0$ with equality if and only if $P(\boldsymbol{y}) \equiv Q(\boldsymbol{y})$, it can be seen as a measure of the dissimilarity (a "distance"[1]) of the distributions $P(\boldsymbol{y})$ and $Q(\boldsymbol{y})$. One can thus think of improving the approximation by minimizing the KL divergence variationally:

[1] Rigorously, it is called "divergence" and not "distance" because of the lack of symmetry, i.e. $\mathrm{KL}(Q\|P) \neq \mathrm{KL}(P\|Q)$).

this motivates our attempt of finding the stationary points of our KL divergence between complex measures

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{\mu}_{\mathrm{gen}}} = 0 \tag{G.1.4a}$$

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{C}_{\mathrm{gen}}} = 0 \tag{G.1.4b}$$

in order to obtain equations for the parameters of the optimal Gaussian $Q(\boldsymbol{y})$. Given the KL expression (G.1.3), the set of equations (G.1.4a) for the components of $\boldsymbol{\mu}_{\mathrm{gen}}$ reduces to

$$\nabla_{\boldsymbol{\mu}_{\mathrm{gen}}} \left(\langle \mathcal{H}(\boldsymbol{y}) + \ln P_0(\boldsymbol{x}) \rangle_Q \right) = 0 \tag{G.1.5}$$

while the equation for the inverse correlation matrix (G.1.4b)

$$\frac{1}{2} \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \ln \left(\det \boldsymbol{C}_{\mathrm{gen}} \right) + \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \langle \mathcal{H}(\boldsymbol{y}) \rangle_Q + \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \langle \ln P_0(\boldsymbol{x}) \rangle_Q = 0 \tag{G.1.6}$$

For the first term in (G.1.6), one can exploit the properties

$$\nabla_{\boldsymbol{C}_{\mathrm{gen}}} \ln \left(\det \boldsymbol{C}_{\mathrm{gen}} \right) = \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \mathrm{Tr}(\ln \boldsymbol{C}_{\mathrm{gen}}) \tag{G.1.7}$$

$$\frac{\partial \, \mathrm{Tr}(\ln \boldsymbol{C}_{\mathrm{gen}})}{\partial (\boldsymbol{C}_{\mathrm{gen}})_{ijtt'}} = (\boldsymbol{C}_{\mathrm{gen}}^{-1})_{jit't} \tag{G.1.8}$$

and the following identities (see [165] for a proof)

$$\nabla_{\boldsymbol{\mu}_{\mathrm{gen}}} \langle \mathcal{H}(\boldsymbol{y}) \rangle_Q = \left\langle \nabla_{\boldsymbol{y}} \mathcal{H}(\boldsymbol{y}) \right\rangle_Q \tag{G.1.9a}$$

$$\nabla_{\boldsymbol{C}_{\mathrm{gen}}} \langle \mathcal{H}(\boldsymbol{y}) \rangle_Q = \frac{1}{2} \left\langle \nabla_{\boldsymbol{y}} \nabla_{\boldsymbol{y}} \mathcal{H}(\boldsymbol{y}) \right\rangle_Q \tag{G.1.9b}$$

To apply these, one has to calculate the derivatives of $\mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ with respect to the variables $\boldsymbol{x}$ and $\mathrm{i}\hat{\boldsymbol{x}}$

$$\frac{\partial \mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{x}})}{\partial (\mathrm{i}\hat{x}_i(t))} = x_i(t + \Delta) - x_i(t) - \Delta \Phi_i(\boldsymbol{x}(t)) + \Delta \sum_j \mathrm{i}\hat{x}_j(t) \Sigma_{ji}(\boldsymbol{x}(t)) \tag{G.1.10}$$

$$\frac{\partial \mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{x}})}{\partial x_i(t)} = \mathrm{i}\hat{x}_i(t - \Delta) - \mathrm{i}\hat{x}_i(t) - \Delta \mathrm{i}\hat{x}_i(t) \frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \frac{\Delta}{2} \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \tag{G.1.11}$$

with

$$\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} = \sum_{j,l,j \neq l} k^+_{ij,l} x_j(t) + \frac{1}{2} \sum_{j,l,j \neq l} k^-_{i,jl} + \sum_j \lambda_{ij} + \sum_l 2 k^+_{ii,l} x_i(t) + \sum_j k^-_{i,jj} \tag{G.1.12}$$

For the second derivatives we need a number of different combinations

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = -\Delta \mathrm{i}\hat{x}_i(t)\frac{\partial}{\partial x_j(t)}\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \frac{\Delta}{2}\sum_{lk}\mathrm{i}\hat{x}_l(t)\frac{\partial}{\partial x_j(t)}\frac{\partial \Sigma_{lk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t) \tag{G.1.13a}$$

$$\frac{\partial}{\partial(\mathrm{i}\hat{x}_i(t))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = -1 - \Delta\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \Delta\sum_k \frac{\partial \Sigma_{ik}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t) \tag{G.1.13b}$$

$$\frac{\partial}{\partial(\mathrm{i}\hat{x}_j(t))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = \Delta\Sigma_{ij}(\boldsymbol{x}(t)) \tag{G.1.13c}$$

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial \mathrm{i}\hat{x}_i(t)} = -\Delta\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} + \Delta\sum_k \mathrm{i}\hat{x}_k(t)\frac{\partial \Sigma_{ki}(\boldsymbol{x}(t))}{\partial x_j(t)} \qquad i \neq j \tag{G.1.13d}$$

$$\frac{\partial}{\partial x_i(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial \mathrm{i}\hat{x}_i(t)} = \frac{\partial}{\partial \mathrm{i}\hat{x}_i(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} \tag{G.1.13e}$$

$$\frac{\partial}{\partial x_i(t+\Delta)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = \frac{\partial}{\partial(\mathrm{i}\hat{x}_i(t-\Delta))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = 1 \tag{G.1.13f}$$

with

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} = \sum_l k^+_{ij,l} \qquad i \neq j \tag{G.1.14a}$$

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} = \sum_l 2k^+_{ii,l} \qquad i = j \tag{G.1.14b}$$

$$\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} = \lambda_{ji} + k^+_{jj,i}x_j(t) + 2k^-_{j,ii} + \sum_l\left(\frac{1}{2}k^+_{jl,i}x_l(t) - k^+_{ij,l}x_i(t)\right) \tag{G.1.14c}$$

For all other time differences the second derivatives are zero.

As a consequence, considering again the $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ variables separately, one can cast the equations (G.1.4a) in the form

$$\left\langle\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} + \frac{\partial(\ln P_0(\boldsymbol{x}))}{\partial x_i(t)}\right\rangle = 0 \qquad \left\langle\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} + \frac{\partial(\ln P_0(\boldsymbol{x}))}{\partial(\mathrm{i}\hat{x}_i(t))}\right\rangle = 0 \tag{G.1.15}$$

$P_0(\boldsymbol{x})$ can be assumed to be Gaussian (since this is the simplest case and it is expected to link up with the variational ansatz we have introduced) and it cannot depend on $\hat{\boldsymbol{x}}$ (since it is just an auxiliary variable), i.e.

$$P_0(\boldsymbol{x}) \equiv P_0(\boldsymbol{x}) \equiv \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_0, \boldsymbol{C}_0) \tag{G.1.16}$$

An explicit calculation immediately shows that the only nonzero derivative is the one for $t = 0$ and in particular

$$\frac{\partial \ln P_0}{\partial x_i(0)} = -\left\{\sum_j (\boldsymbol{C}_0^{-1}(0))_{ij}(\boldsymbol{x}(0) - \boldsymbol{\mu}_0(0))_j\right\} \tag{G.1.17}$$

thus

$$\left\langle\frac{\partial \ln P_0}{\partial x_i(0)}\right\rangle = -\left\{\sum_j (\boldsymbol{C}_0^{-1}(0))_{ij}(\boldsymbol{\mu}(0) - \boldsymbol{\mu}_0(0))_j\right\} \tag{G.1.18}$$

The obvious solution (which we will show later is self-consistent) is that

$$\boldsymbol{\mu}(0) = \boldsymbol{\mu}_0(0) \qquad \boldsymbol{C}_0^{-1}(0) = \boldsymbol{C}^{-1}(0) \tag{G.1.19}$$

With this hypothesis, the average of first derivatives is identically zero, while we have nonzero second derivatives only for $t = 0$ and for the case

$$\left\langle \frac{\partial}{\partial x_j(0)} \frac{\partial \ln P_0(\boldsymbol{x}(0))}{\partial x_i(0)} \right\rangle = -(\boldsymbol{C}^{-1}(0))_{ij} \tag{G.1.20}$$

Then given the derivatives (G.1.10) and (G.1.11) and carrying out the average in (G.1.15) one obtains

$$\mathrm{i}\hat{\mu}_i(t) - \mathrm{i}\hat{\mu}_i(t-\Delta) + \Delta \left\langle \mathrm{i}\hat{x}_i(t) \frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} \right\rangle - \frac{\Delta}{2} \left\langle \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \right\rangle = 0 \tag{G.1.21}$$

while for the ordinary means

$$\mu_i(t+\Delta) - \mu_i(t) - \Delta \langle \Phi_i(\boldsymbol{x}(t)) \rangle + \Delta \left\langle \sum_j \mathrm{i}\hat{x}_j(t) \Sigma_{ji}(\boldsymbol{x}(t)) \right\rangle = 0 \tag{G.1.22}$$

with

$$\begin{aligned}
\langle \Phi_i(\boldsymbol{x}(t)) \rangle = \Phi_i(\boldsymbol{\mu}(t), \boldsymbol{C}(t,t)) = & \sum_{j,l,j\neq l} \left(k^-_{l,ij}\mu_l(t) - k^+_{ij,l}C_{ij}(t,t) - k^+_{ij,l}\mu_i(t)\mu_j(t) \right) + \\
+ \frac{1}{2} & \sum_{j,l,j\neq l} \left(k^+_{jl,i}C_{jl}(t,t) + k^+_{jl,i}\mu_j(t)\mu_l(t) - k^-_{i,jl}\mu_i(t) \right) + \sum_j \left(\lambda_{ji}\mu_j(t) - \lambda_{ij}\mu_i(t) \right) + \\
+ & \sum_l \left(2k^-_{l,ii}\mu_l(t) - k^+_{ii,l}C_{ii}(t,t) - k^+_{ii,l}\mu_i(t)\mu_i(t) \right) + \sum_j \left(\frac{1}{2}k^+_{jj,i}C_{jj}(t,t) + \frac{1}{2}k^+_{jj,i}\mu_j(t)\mu_j(t) - k^-_{i,jj}\mu_i(t) \right)
\end{aligned} \tag{G.1.23}$$

The "initial" (boundary) condition for (G.1.21), at $t = T - \Delta$, which is given by $\langle \partial \mathcal{H}/\partial \boldsymbol{x}(T) \rangle = 0$, yields $\hat{\mu}_i(T - \Delta) = 0$. If we also assume that the approximate solution is causal (i.e. $R_{ij}(t,t) = 0$ as a generic consequence of the Itô discretization [23]), and we solve backwards in time we have that $\hat{\mu}_i(t) = 0\ \forall t$. This is consistent with the fact that averages of $\hat{x}_i(t)$ and all its products should vanish, as a consequence of a normalization requirement $Z[0] = 1$ (as shown in [144]): thus $\hat{\mu}_i(t) \equiv 0\ \forall t$, $-\mathrm{i}\langle \hat{x}_i(t)x_j(t') \rangle \equiv 0$ if $t \geq t'$ and $-\langle \hat{x}_i(t)\hat{x}_j(t') \rangle \equiv 0$, leading also the last term in equation (G.1.22) to zero. Rearranging the latter and taking the limit $\Delta \to 0$ one obtains

$$\frac{d\mu_i(t)}{dt} = \Phi_i(\boldsymbol{\mu}(t), \boldsymbol{C}(t,t)) \sim \Phi_i(\boldsymbol{\mu}(t)) \tag{G.1.24}$$

It represents the ordinary equation of motion for the mean values with additional terms stemming from equal time correlations $\boldsymbol{C}(t,t)$, which nevertheless vanish for large volumes - see section 5.2.

Similarly, equations (G.1.6) can be solved by using the property (G.1.9b) (which involves the calculation of (G.1.13), the second derivatives of $\mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{x}})$); nonzero inverse correlation elements,

for $t \neq 0$, are then given by

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{jitt} = \Delta \mathrm{i}\hat{\mu}_i(t) \sum_l k^+_{ij,l} - \frac{\Delta}{2}\Big\langle \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial}{\partial x_j(t)} \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \Big\rangle \qquad i \neq j \tag{G.1.25a}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{iitt} = \Delta \mathrm{i}\hat{\mu}_i(t) \sum_l 2k^+_{ii,l} - \frac{\Delta}{2}\Big\langle \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial}{\partial x_i(t)} \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \Big\rangle \tag{G.1.25b}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{\hat{\imath}itt} = -1 - \Delta K_{ii}(t) + \Big\langle \Delta \sum_k \frac{\partial \Sigma_{ik}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \Big\rangle \tag{G.1.25c}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{\hat{\imath}\hat{\jmath}tt} = -\Delta \langle \Sigma_{ij}(\boldsymbol{x}(t)) \rangle \tag{G.1.25d}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{j\hat{\imath}tt} = -\Delta K_{ij}(t) + \Big\langle \Delta \sum_k \mathrm{i}\hat{x}_k(t) \frac{\partial \Sigma_{ki}(\boldsymbol{x}(t))}{\partial x_j(t)} \Big\rangle \qquad i \neq j \tag{G.1.25e}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{i\hat{\imath}t+\Delta t} = (\boldsymbol{C}_{\text{gen}}^{-1})_{\hat{\imath}it-\Delta t} = 1 \tag{G.1.25f}$$

where the indexes $\hat{\imath}$ and $\hat{\jmath}$ are introduced to refer to $\hat{x}_i$ and $\hat{x}_j$. We have defined the quantity $K_{ij}(t)$ as follows

$$K_{ij}(t) = \Big\langle \frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \Big\rangle \tag{G.1.26}$$

explicitly

$$K_{ii}(t) = \sum_{jl, j\neq l} k^+_{ij,l}\mu_j(t) + \frac{1}{2}\sum_{jl, j\neq l} k^-_{i,jl} + \sum_j \lambda_{ij} + 2\sum_l k^+_{ii,l}\mu_i(t) + \frac{1}{2}\sum_j k^-_{i,jj} \tag{G.1.27}$$

$$K_{ij}(t) = \lambda_{ji} + k^+_{jj,i}\mu_j(t) + 2k^-_{j,ii} + \sum_l \Big(k^+_{jl,i}\mu_l(t) - k^+_{ij,l}\mu_i(t)\Big) \quad i \neq j \tag{G.1.28}$$

For $t = 0$, the first two equations (G.1.25a) and (G.1.25b) should be slightly modified

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{ji00} = (\boldsymbol{C}_0^{-1})_{ji00} + \Delta\Big[\mathrm{i}\hat{\mu}_i(0) \sum_l k^+_{ij,l} - \frac{1}{2}\Big\langle \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial}{\partial x_j(t)} \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \Big\rangle\Big] \qquad i \neq j \tag{G.1.29a}$$

$$(\boldsymbol{C}_{\text{gen}}^{-1})_{ii00} = (\boldsymbol{C}_0^{-1})_{ii00} + \Delta\Big[2\mathrm{i}\hat{\mu}_i(0) \sum_l k^+_{ii,l} - \frac{1}{2}\Big\langle \sum_{jk} \mathrm{i}\hat{x}_j(t) \frac{\partial}{\partial x_i(t)} \frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)} \mathrm{i}\hat{x}_k(t) \Big\rangle\Big] \tag{G.1.29b}$$

With the solution $\hat{\mu}_i(0) = 0$, eqs. (G.1.29) show that the assumption (G.1.19) is not only reasonable but even necessary.

The aim is then to recover the differential equations for temporal evolution of correlations; let us proceed by imposing the validity of the equality

$$\sum_{jt''} \left[(\boldsymbol{C}_{\text{gen}})_{kjt't''}(\boldsymbol{C}_{\text{gen}})^{-1}_{jit''t} + (\boldsymbol{C}_{\text{gen}})_{k\hat{\jmath}t't''}(\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{\jmath}it''t}\right] = \delta_{ki}\delta_{t't} \tag{G.1.30}$$

where the hatted indices $\hat{\jmath}$ refer to elements containing auxiliary elements $\hat{x}_j$. If we consider explicitly only the nonzero terms, the sum then becomes

$$\begin{aligned}&\sum_{j\neq i}(\boldsymbol{C}_{\text{gen}})_{kjt't}(\boldsymbol{C}_{\text{gen}})^{-1}_{jitt} + (\boldsymbol{C}_{\text{gen}})_{kit't}(\boldsymbol{C}_{\text{gen}})^{-1}_{iitt} + \sum_{j\neq i}(\boldsymbol{C}_{\text{gen}})_{k\hat{\jmath}t't}(\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{\jmath}itt} + \\ &+(\boldsymbol{C}_{\text{gen}})_{k\hat{\imath}t't}(\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{\imath}itt} + (\boldsymbol{C}_{\text{gen}})_{k\hat{\imath}t't-\Delta}(\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{\imath}it-\Delta t} = \delta_{kit't}\end{aligned} \tag{G.1.31}$$

As before, we shall already use the fact that we expect the solution to be causal and all averages of auxiliary variables to vanish. Substituting the expressions (G.1.25), dividing by iΔ and taking the limit for $\Delta \to 0$ yields the differential equations

$$\frac{\partial R_{ki}(t',t)}{\partial t} = -\delta_{ki}\delta(t'-t) - \sum_j R_{kj}(t',t)K_{ji}(t) \qquad t' > t \tag{G.1.32}$$

$$\frac{\partial R_{ki}(t',t)}{\partial t'} = \sum_{j\neq k} K_{kj}(t')R_{ji}(t',t) + \delta_{ki}\delta(t-t') \qquad t' > t \tag{G.1.33}$$

$$\frac{\partial C_{ki}(t',t)}{\partial t} = \sum_j C_{kj}(t',t)K_{ij}(t) + \sum_j R_{kj}(t',t)\langle\Sigma_{ji}(\boldsymbol{x}(t))\rangle \tag{G.1.34}$$

As discussed in the main text, the results of the GVA can be understood as an effective linearization of the Langevin dynamics (5.2.1) around time dependent means. Let us show in more detail that the equations (G.1.32), (G.1.33), (G.1.34) can be mapped into this case. We shall start from the fluctuations about the mean, which obey

$$\frac{d(\boldsymbol{x}(t)-\boldsymbol{\mu}(t))}{dt} = \boldsymbol{K}(t)(\boldsymbol{x}(t)-\boldsymbol{\mu}(t)) + \boldsymbol{\xi}(t) \tag{G.1.35}$$

$\boldsymbol{K}(t)$ represents a time dependent rate matrix (as given by the linearization around time-dependent means (G.1.26)) and the noise satisfies $\langle\boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t')\rangle = \boldsymbol{\Sigma}(\boldsymbol{x}(t))\delta(t-t')$. Simply applying the definition of correlations one has

$$\begin{aligned}\boldsymbol{C}(t',t) = {}& \mathrm{T}\left[e^{\int_0^{t'}\boldsymbol{K}(t'')dt''}\right]\boldsymbol{C}(0,0)\,\mathrm{T}\left[e^{\int_0^{t}\boldsymbol{K}^T(t'')dt''}\right] + \\ &+ \int_0^{\min(t,t')}\theta(t-s)\mathrm{T}\left[e^{\int_s^{t'}\boldsymbol{K}(t'')dt''}\right]\langle\boldsymbol{\Sigma}(\boldsymbol{x}(s))\rangle\,\mathrm{T}\left[e^{\int_s^{t}\boldsymbol{K}^T(t'')dt''}\right]ds\end{aligned} \tag{G.1.36}$$

where the time-ordering operator T has been introduced because time is used as an index to determine the order of $\boldsymbol{K}(t)$ matrices: earlier times on the right and later times on the left.
The following equation can then be derived

$$\frac{\partial\boldsymbol{C}(t',t)}{\partial t} = \theta(t'-t)\mathrm{T}\left[e^{\int_t^{t'}\boldsymbol{K}(t'')dt''}\right]\langle\boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle + \boldsymbol{C}(t',t)\boldsymbol{K}^T(t) \tag{G.1.37}$$

The two-point functions appearing in (G.1.32) and (G.1.33) have the interpretation of response functions

$$\boldsymbol{R}(t',t) = \frac{\partial\langle\boldsymbol{x}(t')\rangle}{\partial\boldsymbol{l}(t)} \tag{G.1.38}$$

where $\boldsymbol{l}(t)$ denotes the external field we can imagine applying to the system at time t, $t < t'$. Its presence leads one to modify the action

$$\mathcal{H} = \int\left(\mathrm{i}\hat{\boldsymbol{x}}(s)\left[\frac{d\boldsymbol{x}(s)}{ds} - \boldsymbol{K}(s)\boldsymbol{x}(s) - \boldsymbol{l}(s) - \boldsymbol{\xi}(s)\right]\right)ds \tag{G.1.39}$$

as well as the general solution

$$\boldsymbol{x}(t) = \boldsymbol{\mu}(t) + \int_0^t \mathrm{T}\left[e^{\int_s^t \boldsymbol{K}(t')dt'}\right](\boldsymbol{l}(s) + \boldsymbol{\xi}(s))\,ds \tag{G.1.40}$$

for $\boldsymbol{x}(0) = 0$. Given this expression, the response (G.1.38) reads explicitly

$$\boldsymbol{R}(t', t) = \theta(t' - t) \cdot \mathrm{T}\left[e^{\int_t^{t'} \boldsymbol{K}(t'')dt''}\right] \tag{G.1.41}$$

whose differential equations are given by

$$\frac{\partial \boldsymbol{R}(t', t)}{\partial t} = -\delta(t' - t)\mathbb{1} - \boldsymbol{R}(t', t)\boldsymbol{K}(t) \tag{G.1.42}$$

$$\frac{\partial \boldsymbol{R}(t', t)}{\partial t'} = \delta(t' - t)\mathbb{1} + \boldsymbol{K}(t')\boldsymbol{R}(t', t) \tag{G.1.43}$$

which are precisely (G.1.32) and (G.1.33). By substitution of equation (G.1.41) in equation (G.1.37), one obtains

$$\frac{\partial \boldsymbol{C}(t', t)}{\partial t} = \boldsymbol{C}(t', t)\boldsymbol{K}^T(t) + \boldsymbol{R}(t', t)\langle\boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \tag{G.1.44}$$

which is the vectorial version of (G.1.34) and for equal times gives

$$\frac{\partial \boldsymbol{C}(t, t)}{\partial t} = \boldsymbol{C}(t, t)\boldsymbol{K}^T(t) + \boldsymbol{K}(t)\boldsymbol{C}(t, t) + \langle\boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \tag{G.1.45}$$

Appendix H

Memory and coloured noise in the linearized dynamics

H.1 Stationary case

The Gaussian variational approach effectively leads to a linearization of the dynamics and the structure of the subnetwork reduced dynamics can be worked out by matrix inversion. To show this, we consider for simplicity a dynamics linearized around *steady state* means and then generalize the expressions for the memory and the effective noise to the case of time-dependent means. We thus start from

$$\boldsymbol{x} = \boldsymbol{\mu} + \delta\boldsymbol{x} \tag{H.1.1}$$

$$\frac{d\delta\boldsymbol{x}}{dt} = \boldsymbol{K}\delta\boldsymbol{x} + \boldsymbol{\xi}(t) \tag{H.1.2}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{K}$ are constant as calculated at steady state. To derive the memory, i.e. the convolution of a memory kernel and past subnetwork trajectories, it is convenient to work in Fourier space, where convolutions in time become algebraic products. Let us focus on the subnetwork-subnetwork generalized covariance

$$(\tilde{\boldsymbol{C}}^{\mathrm{ss}}_{\mathrm{gen}}) = \begin{pmatrix} \tilde{\boldsymbol{C}}^{\mathrm{ss}} & \tilde{\boldsymbol{R}}^{\mathrm{ss}} \\ (\tilde{\boldsymbol{R}}^{\mathrm{ss}})^T & \mathbb{0} \end{pmatrix}$$

For the sake of brevity we dropped the frequency dependence, i.e. $\tilde{\boldsymbol{C}}^{\mathrm{ss}}_{\mathrm{gen}} = \tilde{\boldsymbol{C}}^{\mathrm{ss}}_{\mathrm{gen}}(\omega)$.
We are interested in the inverse of $\tilde{\boldsymbol{C}}^{\mathrm{ss}}_{\mathrm{gen}}$ as it describes the marginal dynamics for subnetwork variables in the variational approximation; we thus expect it to have the structure

$$(\tilde{\boldsymbol{C}}^{\mathrm{ss}}_{\mathrm{gen}})^{-1} = \begin{pmatrix} \mathbb{0} & -\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T} \\ -\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}\,T} + \tilde{\boldsymbol{M}}^{\mathrm{ss}} & \tilde{\boldsymbol{N}}^{\mathrm{ss}}_0 \end{pmatrix}$$

The first term of off diagonal terms $-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}}$ reproduces the internal subnetwork dynamics, $\tilde{\boldsymbol{M}}^{\mathrm{ss}}$ is the memory function (the transposed T is due to how the order of indices is set in equation (5.4.2)) while $\tilde{\boldsymbol{N}}_0^{\mathrm{ss}}$ is the covariance of an effective coloured noise. By means of the simple identity

$$(\tilde{\boldsymbol{C}}_{\mathrm{gen}}^{\mathrm{ss}})^{-1}\tilde{\boldsymbol{C}}_{\mathrm{gen}}^{\mathrm{ss}} = \mathbb{1} \tag{H.1.3}$$

the following relations can be straightforwardly retrieved

$$(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T})\tilde{\boldsymbol{R}}^{\mathrm{ss}} = \mathbb{1} \tag{H.1.4}$$

$$(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}\,T} + \tilde{\boldsymbol{M}}^{\mathrm{ss}})\tilde{\boldsymbol{C}}^{\mathrm{ss}} + \tilde{\boldsymbol{N}}_0^{\mathrm{ss}}\tilde{\boldsymbol{R}}^{\mathrm{ss}\,T} = \mathbb{0} \tag{H.1.5}$$

One can thus achieve the marginalization over bulk variables by simple matrix inversion; to find $\tilde{\boldsymbol{M}}^{\mathrm{ss}}$ and $\tilde{\boldsymbol{N}}_0^{\mathrm{ss}}$, we need the explicit expressions for $\tilde{\boldsymbol{R}}^{\mathrm{ss}}$. The response function is defined as the Green's function associated with (H.1.2)

$$\left[\frac{\partial}{\partial t} - \boldsymbol{K}\right]\boldsymbol{R}(t,t') = \delta(t-t') \tag{H.1.6}$$

and simply given by

$$\boldsymbol{R}(t,t') = \left[\frac{\partial}{\partial t} - \boldsymbol{K}\right]^{-1}\delta(t-t') \tag{H.1.7}$$

In Fourier representation the response function becomes

$$\tilde{\boldsymbol{R}}(\omega) = [-\mathrm{i}\omega - \boldsymbol{K}]^{-1} \tag{H.1.8}$$

and the inverse immediately follows

$$(\tilde{\boldsymbol{R}})^{-1}(\omega) = [-\mathrm{i}\omega - \boldsymbol{K}] \tag{H.1.9}$$

In the blocks form of matrices introduced in the main text they read

$$\tilde{\boldsymbol{R}} = \begin{pmatrix} \tilde{\boldsymbol{R}}^{\mathrm{ss}} & \tilde{\boldsymbol{R}}^{\mathrm{sb}} \\ \tilde{\boldsymbol{R}}^{\mathrm{bs}} & \tilde{\boldsymbol{R}}^{\mathrm{bb}} \end{pmatrix}$$

$$(\tilde{\boldsymbol{R}})^{-1} = \begin{pmatrix} \tilde{\boldsymbol{R}}^{\mathrm{ss}} & \tilde{\boldsymbol{R}}^{\mathrm{sb}} \\ \tilde{\boldsymbol{R}}^{\mathrm{bs}} & \tilde{\boldsymbol{R}}^{\mathrm{bb}} \end{pmatrix}^{-1} = \begin{pmatrix} -\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} & \boldsymbol{K}^{\mathrm{sb}} \\ \boldsymbol{K}^{\mathrm{bs}} & -\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{bb}} \end{pmatrix}$$

Given a matrix in blocks form

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}$$

to calculate the blocks of the inverse

$$\boldsymbol{Q}^{-1} = \begin{pmatrix} \boldsymbol{A}' & \boldsymbol{B}' \\ \boldsymbol{C}' & \boldsymbol{D}' \end{pmatrix}$$

one can use standard formulas [16]

$$\boldsymbol{A}' = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \tag{H.1.10a}$$

$$\boldsymbol{B}' = -(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \tag{H.1.10b}$$

$$\boldsymbol{C}' = -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} \tag{H.1.10c}$$

$$\boldsymbol{D}' = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \tag{H.1.10d}$$

Equation (H.1.4) implies for the memory function $\tilde{\boldsymbol{M}}^{\mathrm{ss}\,T}$

$$-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T} = (\tilde{\boldsymbol{R}}^{\mathrm{ss}})^{-1} \tag{H.1.11}$$

By using formula (H.1.10a) one immediately obtains an expression for $(\tilde{\boldsymbol{R}}^{\mathrm{ss}})^{-1}$, therefore

$$\tilde{\boldsymbol{M}}^{\mathrm{ss}\,T}(\omega) = \boldsymbol{K}^{\mathrm{sb}}(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{bb}})^{-1}\boldsymbol{K}^{\mathrm{bs}} \tag{H.1.12}$$

which in the time domain translates into

$$(\boldsymbol{M}^{\mathrm{ss}})^{T}(t,t') = \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}} \tag{H.1.13}$$

Moreover, equation (H.1.5) leads to an expression for the "effective" noise in the subnetwork $\tilde{\boldsymbol{N}}_0^{\mathrm{ss}}$

$$\tilde{\boldsymbol{N}}_0^{\mathrm{ss}} = -(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T})\tilde{\boldsymbol{C}}^{\mathrm{ss}}(\tilde{\boldsymbol{R}}^{\mathrm{ss}\,T})^{-1} = -(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T})\tilde{\boldsymbol{C}}^{\mathrm{ss}}(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}\,T} + \tilde{\boldsymbol{M}}^{\mathrm{ss}}) \tag{H.1.14}$$

From (G.1.36) one can write

$$\tilde{\boldsymbol{C}} = (-\mathrm{i}\omega + \boldsymbol{K})^{-1}(\boldsymbol{C}(0,0) + \boldsymbol{\Sigma})(-\mathrm{i}\omega + \boldsymbol{K}^{T})^{-1} = \tilde{\boldsymbol{R}}\,(\boldsymbol{C}(0,0) + \boldsymbol{\Sigma})\,\tilde{\boldsymbol{R}}^{T} \tag{H.1.15}$$

For simplicity we can set the off-diagonal blocks to zero (i.e. $\boldsymbol{C}^{\mathrm{sb}}(0,0) = \boldsymbol{C}^{\mathrm{bs}}(0,0) = \boldsymbol{\Sigma}^{\mathrm{sb}} = \boldsymbol{\Sigma}^{\mathrm{bs}} = \mathbb{0}$), thus (H.1.15) gives

$$\tilde{\boldsymbol{C}}^{\mathrm{ss}} = \tilde{\boldsymbol{R}}^{\mathrm{ss}}\,(\boldsymbol{C}^{\mathrm{ss}}(0,0) + \boldsymbol{\Sigma}^{\mathrm{ss}})\,\tilde{\boldsymbol{R}}^{\mathrm{ss}\,T} + \tilde{\boldsymbol{R}}^{\mathrm{sb}}\,(\boldsymbol{C}^{\mathrm{bb}}(0,0) + \boldsymbol{\Sigma}^{\mathrm{bb}})\,\tilde{\boldsymbol{R}}^{\mathrm{bs}\,T} \tag{H.1.16}$$

By applying the block inversion formula (H.1.10) and using (H.1.12) we find

$$\tilde{\boldsymbol{R}}^{\mathrm{ss}} = (-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T})^{-1} \tag{H.1.17a}$$

$$\tilde{\boldsymbol{R}}^{\mathrm{sb}} = (-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{ss}} + \tilde{\boldsymbol{M}}^{\mathrm{ss}\,T})^{-1}\boldsymbol{K}^{\mathrm{sb}}(-\mathrm{i}\omega - \boldsymbol{K}^{\mathrm{bb}})^{-1} \tag{H.1.17b}$$

We insert (H.1.16) and (H.1.17) in (H.1.14) and we rewrite it in the time domain

$$\begin{aligned}
\boldsymbol{N}_0^{\mathrm{ss}}(t,t') =& \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{\Sigma}^{\mathrm{ss}}\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}\theta(t)e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)\theta(t')e^{(\boldsymbol{K}^{\mathrm{bb}})^{T}t'}(\boldsymbol{K}^{\mathrm{sb}})^{T} \\
&+ \boldsymbol{K}^{\mathrm{sb}}\int dt''\theta(t-t'')e^{\boldsymbol{K}^{\mathrm{bb}}(t-t'')}\boldsymbol{\Sigma}^{\mathrm{bb}}\theta(t'-t'')e^{(\boldsymbol{K}^{\mathrm{bb}})^{T}(t'-t'')}(\boldsymbol{K}^{\mathrm{sb}})^{T} = \\
&= \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{\Sigma}^{\mathrm{ss}}\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)e^{(\boldsymbol{K}^{\mathrm{bb}})^{T}t'}(\boldsymbol{K}^{\mathrm{sb}})^{T} \\
&+ \boldsymbol{K}^{\mathrm{sb}}\int_0^{\min(t,t')} dt''e^{\boldsymbol{K}^{\mathrm{bb}}(t-t'')}\boldsymbol{\Sigma}^{\mathrm{bb}}e^{(\boldsymbol{K}^{\mathrm{bb}})^{T}(t'-t'')}(\boldsymbol{K}^{\mathrm{sb}})^{T}
\end{aligned} \tag{H.1.18}$$

Note that the initial term comes from imposing the initial condition in the form

$$\partial_t \boldsymbol{x}^{\mathrm{s}}(t) = \boldsymbol{K}^{\mathrm{ss}}\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{x}^{\mathrm{b}}(t) + \boldsymbol{\xi}^{\mathrm{s}}(t) + \boldsymbol{x}^{\mathrm{s}}(0)\delta(t) \tag{H.1.19}$$

As clear from (H.1.16), it is crucial to retain off diagonal blocks as $\tilde{\boldsymbol{R}}^{\mathrm{sb}}$, $\tilde{\boldsymbol{R}}^{\mathrm{bs}}$ to properly capture the propagation of fluctuations through the bulk.

H.2 Generalization to non-stationary case

Let us assume a linearization around time-dependent means, i.e. let us take the rate matrix as a function of time $\boldsymbol{K}(t)$ as well as the noise covariance $\langle \boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle$. We can generalize (H.1.13) to this case and we get

$$(\boldsymbol{M}^{\mathrm{ss}})^T(t,t') = \boldsymbol{K}^{\mathrm{sb}}(t)\,\mathrm{T}\left[e^{\int_{t'}^{t}\boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\boldsymbol{K}^{\mathrm{bs}}(t') \tag{H.2.1}$$

Similarly for (H.1.18) one has

$$\begin{aligned}\boldsymbol{N}_0^{\mathrm{ss}}(t,t') &= \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{K}^{\mathrm{sb}}(t)\,\mathrm{T}\left[e^{\int_0^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\boldsymbol{C}^{\mathrm{bb}}(0,0)\,\mathrm{T}\left[e^{\int_0^{t'}(\boldsymbol{K}^{\mathrm{bb}}(s))^T ds}\right](\boldsymbol{K}^{\mathrm{sb}})^T(t') \\ &+ \langle\boldsymbol{\Sigma}^{\mathrm{ss}}(\boldsymbol{x}(t))\rangle\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}(t)\int_0^{\min(t,t')} dt''\,\mathrm{T}\left[e^{\int_{t''}^{t}\boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\langle\boldsymbol{\Sigma}^{\mathrm{bb}}(\boldsymbol{x}(t))\rangle\,\mathrm{T}\left[e^{\int_{t''}^{t'}(\boldsymbol{K}^{\mathrm{bb}})^T(s)ds}\right](\boldsymbol{K}^{\mathrm{sb}})^T(t'')\end{aligned} \tag{H.2.2}$$

which can be rewritten

$$\begin{aligned}\boldsymbol{N}_0^{\mathrm{ss}}(t,t') &= \boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t') + \boldsymbol{K}^{\mathrm{sb}}(t)\boldsymbol{R}^{\mathrm{bb}}(t)\boldsymbol{C}^{\mathrm{bb}}(0,0)\boldsymbol{R}^{\mathrm{bb}\,T}(t')(\boldsymbol{K}^{\mathrm{sb}})^T(t') \\ &+ \langle\boldsymbol{\Sigma}^{\mathrm{ss}}(\boldsymbol{x}(t))\rangle\delta(t-t') + \boldsymbol{K}^{\mathrm{sb}}(t)\int_0^{\min(t,t')} dt''\boldsymbol{R}^{\mathrm{bb}}(t,t'')\langle\boldsymbol{\Sigma}^{\mathrm{bb}}(\boldsymbol{x}(t))\rangle\boldsymbol{R}^{\mathrm{bb}\,T}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T(t'')\end{aligned} \tag{H.2.3}$$

where we used the definition (G.1.41) for the marginal bulk responses $\boldsymbol{R}^{\mathrm{bb}}(t,t'')$.

Appendix I

Nonlinear corrections by perturbation theory

I.1 Conditional Gaussian moments

Let us here summarize how to find the moments of the Gaussian conditional distribution we denoted as $Q_0(\boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}}|\boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}})$. We recall that the marginal auxiliary means $\hat{\boldsymbol{\mu}}^{\mathrm{s}} = \hat{\boldsymbol{\mu}}^{\mathrm{b}} = 0$ and we assume that the variables are shifted, thus also the marginal ordinary means are zero ($\boldsymbol{\mu}^{\mathrm{s}} = \boldsymbol{\mu}^{\mathrm{b}} = 0$). Conditioning on the set of subnetwork variables $\boldsymbol{x}^{\mathrm{s}}$ and $\hat{\boldsymbol{x}}^{\mathrm{s}}$, the resulting bulk conditional means are such that the full Gaussian can be expressed as a Gaussian only over bulk variables, thus describing the isolated bulk evolution

$$\begin{aligned}
\mathcal{H} = \int dt \Big\{ & (\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t))^T (\partial_t \boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t) - \boldsymbol{K}^{\mathrm{ss}} \boldsymbol{x}^{\mathrm{s}}(t)) - \frac{1}{2} (\hat{\boldsymbol{x}}^{\mathrm{s}}(t))^T \boldsymbol{\Sigma}^{\mathrm{ss}} \hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \\
& + (\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t))^T (\partial_t \boldsymbol{x}^{\mathrm{b}}(t) - \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{x}^{\mathrm{b}}(t)) - \frac{1}{2} (\hat{\boldsymbol{x}}^{\mathrm{b}}(t))^T \boldsymbol{\Sigma}^{\mathrm{bb}} \hat{\boldsymbol{x}}^{\mathrm{b}}(t) \Big\} - \frac{1}{2} (\boldsymbol{x}^{\mathrm{s}}(0))^T \boldsymbol{C}^{\mathrm{ss}}(0,0)^{-1} \boldsymbol{x}^{\mathrm{s}}(0) + \\
& - \frac{1}{2} (\boldsymbol{x}^{\mathrm{b}}(0))^T \boldsymbol{C}^{\mathrm{bb}}(0,0)^{-1} \boldsymbol{x}^{\mathrm{b}}(0) \equiv \text{const.} + \int dt \Big\{ (\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) - \hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t))^T (\partial_t - \boldsymbol{K}^{\mathrm{bb}}) (\boldsymbol{x}^{\mathrm{b}}(t) - \boldsymbol{\mu}^{\mathrm{b|s}}(t)) + \\
& + \frac{1}{2} (\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) - \hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t))^T \boldsymbol{\Sigma}^{\mathrm{bb}} (\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) - \hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)) \Big\} - \frac{1}{2} (\boldsymbol{x}^{\mathrm{b}}(0) - \boldsymbol{\mu}^{\mathrm{b|s}}(0))^T \boldsymbol{C}^{\mathrm{bb}}(0,0)^{-1} (\boldsymbol{x}^{\mathrm{b}}(0) - \boldsymbol{\mu}^{\mathrm{b|s}}(0))
\end{aligned} \tag{I.1.1}$$

where the second term in the r.h.s. can be rewritten as

$$-\frac{1}{2} \int dt \, (\boldsymbol{y}^{\mathrm{b}}(t) - \boldsymbol{\mu}^{\mathrm{b|s}}_{\mathrm{gen}}(t))^T (\boldsymbol{C}^{\mathrm{bb|s}}_{\mathrm{gen}})^{-1}(t,t') (\boldsymbol{y}^{\mathrm{b}}(t) - \boldsymbol{\mu}^{\mathrm{b|s}}_{\mathrm{gen}}(t)) \tag{I.1.2}$$

and

$$(\boldsymbol{C}^{\mathrm{bb|s}}_{\mathrm{gen}})^{-1}(t,t') = \begin{pmatrix} \mathbb{0} & (\partial_t - \boldsymbol{K}^{\mathrm{bb}}) \delta(t-t') \\ (\partial_t - \boldsymbol{K}^{\mathrm{bb}})^T \delta(t-t') & \boldsymbol{\Sigma}^{\mathrm{bb}} \delta(t-t') \end{pmatrix} \tag{I.1.3}$$

with $t, t' \neq 0$. The solution for $\boldsymbol{C}^{\mathrm{bb|s}}(t,t')$ is then derived from the top left block of the inverse $\boldsymbol{C}^{\mathrm{bb|s}}_{\mathrm{gen}}$, to which initial conditions $\boldsymbol{C}^{\mathrm{bb}}(0,0)$ must be added to obtain

$$\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') = \int_0^{\min(t,t'')} dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} \boldsymbol{\Sigma}^{\mathrm{bb}} e^{(\boldsymbol{K}^{\mathrm{bb}})^T(t''-t')} + e^{\boldsymbol{K}^{\mathrm{bb}}t} \boldsymbol{C}^{\mathrm{bb}}(0,0) e^{(\boldsymbol{K}^{\mathrm{bb}})^T t''} \tag{I.1.4}$$

For the means, we set the equality between the l.h.s. and r.h.s. of (I.1.1) for the factors linear in $\hat{\boldsymbol{x}}^{\mathrm{b}}(t)$ and $\boldsymbol{x}^{\mathrm{b}}(t)$

$$-\int dt\, \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)^T \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t) = -\int dt \Big\{ \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)^T (\partial_t - \boldsymbol{K}^{\mathrm{bb}}) \boldsymbol{x}^{\mathrm{b}}(t) \Big\} - \boldsymbol{\mu}^{\mathrm{b|s}}(0)^T \boldsymbol{C}^{\mathrm{bb}}(0,0)^{-1} \boldsymbol{x}^{\mathrm{b}}(0) \tag{I.1.5}$$

Integration by part from 0 to a final time T

$$\begin{aligned} &\int dt\, \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)^T \partial_t \boldsymbol{x}^{\mathrm{b}}(t) = \int dt [\partial_t (\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)^T \boldsymbol{x}^{\mathrm{b}}(t)) - \partial_t \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)^T \boldsymbol{x}^{\mathrm{b}}(t)] = \\ &= \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(T)^T \boldsymbol{x}^{\mathrm{b}}(T) - \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0)^T \boldsymbol{x}^{\mathrm{b}}(0) - \int dt\, \partial_t (\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)^T) \boldsymbol{x}^{\mathrm{b}}(t) \end{aligned} \tag{I.1.6}$$

By an argument analogous to the one for marginal means, one can show that the final condition $\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(T) = 0$ while the terms for $t = 0$ must compensate

$$-\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0)^T \boldsymbol{x}^{\mathrm{b}}(0) - \boldsymbol{\mu}^{\mathrm{b|s}}(0)^T \boldsymbol{C}^{\mathrm{bb}}(0,0)^{-1} \boldsymbol{x}^{\mathrm{b}}(0) = 0 \tag{I.1.7}$$

$$\boldsymbol{\mu}^{\mathrm{b|s}}(0) = -\boldsymbol{C}^{\mathrm{bb}}(0,0) \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0) \tag{I.1.8}$$

which gives a connection between initial conditions. The rest of the equality leads to the temporal evolution, valid $\forall t$

$$\partial_t (\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)) = -\boldsymbol{K}^{\mathrm{sb}\,T} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{bb}\,T} \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) \tag{I.1.9}$$

solved by

$$\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) = \int_t^T dt' e^{-(\boldsymbol{K}^{\mathrm{bb}})^T (t-t')} (\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{I.1.10}$$

which provides, via (I.1.8), an initial condition for $\boldsymbol{\mu}^{\mathrm{b|s}}(t)$. The conditional mean for hat variables is not identically zero as $\hat{\boldsymbol{x}}^{\mathrm{s}}$ must be treated as a *fixed* variable, needed to find the subnetwork reduced dynamics. Analogously, the terms multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t)^T$ lead to the differential equation

$$\partial_t \boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{\mu}^{\mathrm{b|s}}(t) - \boldsymbol{\Sigma}^{\mathrm{bb}} \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) \tag{I.1.11}$$

The explicit solution of this evolves in time as

$$\begin{aligned}
\boldsymbol{\mu}^{\mathrm{b|s}}(t) &= \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}[\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t') - \boldsymbol{\Sigma}^{\mathrm{bb}}\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t')] + e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{\mu}^{\mathrm{b|s}}(0) = \\
&= \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}[\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t') - \boldsymbol{\Sigma}^{\mathrm{bb}}\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t')] - e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(0) = \\
&= \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}[\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t')] - \int_0^t dt' \int_{t'}^T dt'' e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t)}\boldsymbol{\Sigma}^{\mathrm{bb}} e^{(\boldsymbol{K}^{\mathrm{bb}})^T(t'-t'')}(\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'') \\
&\quad - e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)\int_0^T dt' e^{-(\boldsymbol{K}^{\mathrm{bb}})^T t'}(\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') = \qquad \text{(I.1.12)} \\
&= \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt''\Big[\int_0^{\min(t,t'')} dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{\Sigma}^{\mathrm{bb}} e^{(\boldsymbol{K}^{\mathrm{bb}})^T(t''-t')} + \\
&\quad + e^{\boldsymbol{K}^{\mathrm{bb}}t}\boldsymbol{C}^{\mathrm{bb}}(0,0)e^{(\boldsymbol{K}^{\mathrm{bb}})^T t''}\Big](\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'') \\
&= \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt''\boldsymbol{C}^{\mathrm{bb|s}}(t,t'')(\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'')
\end{aligned}$$

The fourth line includes a term reflecting simply the equation of motion and two additional pieces with the temporal propagation of bulk initial conditions and bulk intrinsic noise: as shown in the last line, these two pieces combine to form the conditional bulk correlator (I.1.4). As a general result, Gaussian conditional means (I.1.12) and (I.1.10) are linearly related to the variables on which one conditions while connected conditional correlations (I.1.4) do not depend on their particular values (they depend on the procedure of conditioning as such but not on the particular conditions). For nonzero marginal means (I.1.12) becomes

$$\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\mu}^{\mathrm{b}} + \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{bs}}\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt''\boldsymbol{C}^{\mathrm{bb|s}}(t,t'')(\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'') \qquad \text{(I.1.13)}$$

Finally the conditional correlator of auxiliary variables $\boldsymbol{B}^{\mathrm{bb|s}}(t,t'') \equiv 0$ and the equal times conditional response $\boldsymbol{R}^{\mathrm{bb|s}}(t,t) \equiv 0$, as would be also clear from the block inversion in discrete time of (I.1.3).

I.1.1 Comparison with the Extended Plefka Expansion and the Kalman Filter

The expressions for the conditional bulk means as a function of time found in section I.1 are the same as for the Extended Plefka Expansion with hidden nodes (chapter 3), which is constructed as well as a Gaussian representation of the dynamics although with slight differences. There, as can be seen from the comparison of the inverse generalized conditional covariances (3.2.33) and (I.1.3), working with the effective fields $\hat{C}_i^{\mathrm{eff}}(t,t')$ and $\hat{B}_i^{\mathrm{eff}}(t,t')$ couples the variances of x_i and $\hat{x}_i$, in such a way that e.g. $\boldsymbol{B}^{\mathrm{bb|s}}(t,t'')$ is not identically zero. In addition, the main quantity in chapter 3 is the data likelihood $P(\boldsymbol{x}^{\mathrm{s}})$ (see (3.2.24)), obtained by conditioning on a particular sequence of observations $\boldsymbol{x}^{\mathrm{s}}$ and given by a path integral over $\boldsymbol{x}^{\mathrm{b}}$, $\hat{\boldsymbol{x}}^{\mathrm{b}}$ and $\hat{\boldsymbol{x}}^{\mathrm{s}}$ of the joint probability

$P(\boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}^{\mathrm{s}})$. Here the aim is to find the marginal probability of subnetwork trajectories, which requires to condition jointly on $\boldsymbol{x}^{\mathrm{s}}$ and $\hat{\boldsymbol{x}}^{\mathrm{s}}$ (as their combination encodes the subnetwork dynamics) and is obtained via integration over $\boldsymbol{x}^{\mathrm{b}}$ and $\hat{\boldsymbol{x}}^{\mathrm{b}}$ of $P(\boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}^{\mathrm{s}})$ (see (5.5.1)). Therefore, when comparing the respective moments, the only caveat is that the role of $\hat{\mu}_a$ in chapter 3 is played here by $\hat{x}_a$, which must be kept as a non-averaged variable given that in (3.2.24) there is an integration over $\hat{\boldsymbol{x}}^{\mathrm{s}}$ which does not appear in the GVA effective action (5.5.1). Bearing this in mind, (3.2.27) corresponds to (I.1.11) while (3.2.28) to (I.1.9) (without the assumption of a diagonal noise covariance). These expressions for the means of a stochastic linear dynamics are exact, as one can immediately check by comparison with the Kalman filter and smoother. What we write here is not the standard Kalman filter recursion for the posterior mean but we re-express it in a way that allows us to show its equivalence to (I.1.11) and (I.1.9). As explained in appendix E, the conditional bulk mean computed time step by time step via this algorithm can be obtained from the average of

$$P(\boldsymbol{x}^{\mathrm{b}}(t+\Delta), \boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{X}^{\mathrm{s}}) = \hat{\beta}_{t+\Delta} P(\boldsymbol{x}^{\mathrm{b}}(t+\Delta)|\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) P(\boldsymbol{x}^{\mathrm{s}}(t+\Delta)|\boldsymbol{x}^{\mathrm{b}}(t), \boldsymbol{x}^{\mathrm{s}}(t)) \tag{I.1.14}$$

which can be written explicitly by using the definitions (E.1.3) and (E.1.4). $\hat{\beta}_t$ (see (E.1.8)) is the "backward" variable that together with the forward one $\hat{\alpha}_t$ (see (E.1.7)) allows one to obtain the posterior distribution (E.1.9), i.e.

$$P(\boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{X}^{\mathrm{s}}) = \hat{\alpha}_t \hat{\beta}_t = \mathcal{N}(\boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t), \boldsymbol{C}^{\mathrm{bb}|\mathrm{s}}(t,t)) \tag{I.1.15}$$

By appeal to the MSRJD formalism, one can express

$$\begin{aligned} &P(\boldsymbol{x}^{\mathrm{b}}(t+\Delta), \boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{X}^{\mathrm{s}}) \sim \\ &\int d\hat{\boldsymbol{x}}^{\mathrm{b}}(t) d\hat{\boldsymbol{x}}^{\mathrm{s}}(t) \hat{\beta}_{t+\Delta} e^{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,T}(t)(\boldsymbol{x}^{\mathrm{b}}(t+\Delta) - \boldsymbol{x}^{\mathrm{b}}(t) - \Delta \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{x}^{\mathrm{b}}(t) - \Delta \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{\mathrm{s}}(t))} e^{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,T}(t) \Delta \boldsymbol{\Sigma}^{\mathrm{bb}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t)/2} \\ &\qquad e^{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)(\boldsymbol{x}^{\mathrm{s}}(t+\Delta) - \boldsymbol{x}^{\mathrm{s}}(t) - \Delta \boldsymbol{K}^{\mathrm{ss}} \boldsymbol{x}^{\mathrm{s}}(t) - \Delta \boldsymbol{K}^{\mathrm{sb}} \boldsymbol{x}^{\mathrm{b}}(t))} e^{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t) \Delta \boldsymbol{\Sigma}^{\mathrm{ss}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)/2} \end{aligned} \tag{I.1.16}$$

Deriving the exponent w.r.t. $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t)$ and setting this derivative equal to zero allows one to find the mean of $\boldsymbol{x}^{\mathrm{b}}(t+\Delta)$ conditioned on $\boldsymbol{x}^{\mathrm{b}}(t)$ and $\boldsymbol{X}^{\mathrm{s}}$; this encodes the dynamics

$$\frac{\boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t+\Delta) - \boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t)}{\Delta} = \boldsymbol{K}^{\mathrm{bb}} \boldsymbol{\mu}^{\mathrm{b}|\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{bs}} \boldsymbol{x}^{s}(t) - \boldsymbol{\Sigma}^{\mathrm{bb}} \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t) \tag{I.1.17}$$

which is the discrete time version of (I.1.11). It depends on the values of $\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t)$, whose dynamics follows the stationarity condition of the exponent of (I.1.16) w.r.t. $\boldsymbol{x}^{\mathrm{b}}(t)$, i.e.

$$\frac{\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t) - \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t-\Delta)}{\Delta} = -\boldsymbol{K}^{\mathrm{bb}\,T} \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b}|\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}\,T} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) \tag{I.1.18}$$

which is precisely (I.1.9) in discrete time steps. In the inference context, one also needs the average evolution of $\hat{\boldsymbol{x}}^{\mathrm{s}}(t)$, which is connected to the subnetwork (observed) trajectory $\boldsymbol{X}^{\mathrm{s}}$ (see chapter 3).

As a conclusion, the GVA, as well as the Extended Plefka Expansion with hidden nodes (chapter 3), estimate *exactly* the dynamics of conditional means from the average of $P(\boldsymbol{x}^{\mathrm{b}}(t+\Delta), \boldsymbol{x}^{\mathrm{b}}(t)|\boldsymbol{X}^{\mathrm{s}})$, which is also what the Kalman filter (and smoother) gives for the posterior bulk (hidden) dynamics.

The comparison with the extended Plefka expansion can be developed further. It is actually worth stressing that, for the linearized setting, the mean field version of the GVA effective subnetwork dynamics matches the one encoded by the data likelihood defined in chapter 3, where the observed nodes play the role of the subnetwork we want to reduce our description to. The data likelihood written as a path integral (see (3.2.24) and (3.2.25)) expresses the reduced dynamics of observations as obtained via a marginalization of the hidden (bulk) trajectories; one can deduce it by inspection of (3.2.25) and in the $\Delta \to 0$ limit one has

$$\partial_t x_a(t) = -\lambda x_a(t) + I_a^{\mathrm{eff}}(t) + \chi_a(t) \tag{I.1.19}$$

with $\langle \chi_a(t)\chi_a(t')\rangle = \Sigma_a \delta(t-t') + \hat{B}_a^{\mathrm{eff}}(t,t')$ (we have already included the intrinsic noise contribution in this effective noise). The effective fields $I_a^{\mathrm{eff}}(t)$ and $\hat{B}_a^{\mathrm{eff}}(t,t')$ are given respectively by (3.2.23b) and (3.2.23g), where for simplicity we can set the marginal means to zero. To obtain the marginal subnetwork dynamics, one needs to substitute then the solution for conditional bulk means and correlations (given by (I.1.12) and (I.1.4)) and $\hat{x}_a(t)$ in (I.1.12) should be understood in this comparison as a conditional mean for the reasons we previously pointed out. In this way (I.1.19) becomes

$$\partial_t x_a(t) = -\lambda x_a(t) + \sum_b K_{ab} x_b(t) + \sum_j \int_0^t dt' K_{aj} K_{ja} R_j(t-t') x_a(t') + \chi_a(t) \tag{I.1.20}$$

with $\langle \chi_a(t)\chi_a(t')\rangle = \Sigma_a \delta(t-t') + K_{aj}^2 [R_j(t,0)C_j(0,0)R_j(t',0) + \int_0^{\min(t,t')} dt'' R_j(t,t'')\Sigma_j R_j(t',t'')]$ and $R_j(t-t') = \left[e^{(\boldsymbol{K}^{\mathrm{bb}})(t-t')}\right]_{jj}$ (we bear in mind that the Plefka second moments are purely diagonal). One clearly sees that equation (I.1.20) is the single site analogue of the linear effective subnetwork dynamics (5.4.2), with memory and effective noise covariance given by (5.4.6) and (5.4.7) (for simplicity we considered $C_a(0,0) = 0$).

I.2 Perturbative expansion

The effective action can be treated as a field theory free energy and can be expressed perturbatively [25]. This corresponds, in some respects, to thinking about the subnetwork as a magnetic "field" that forces the mean value of the "order parameter" (i.e. a bulk concentration) to assume a value different from zero: thus the problem is substantially that of calculating a free energy (the reduced

action) for a system in a magnetic field. Let us decompose the action into a "non-interacting" part (containing only terms quadratic in all the variables) and an interacting one (containing higher powers), which can be seen as a small perturbation

$$\mathcal{H} = \mathcal{H}_0 + \Delta\mathcal{H} \tag{I.2.1}$$

The expression for the effective action (5.5.1), applying the idea of a perturbative expansion, becomes

$$\begin{aligned}
\mathcal{H}_{\text{eff}} &= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{(\mathcal{H}_0+\Delta\mathcal{H})} = \\
&= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} \left(1 + \frac{\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}}{\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}} \Delta\mathcal{H} + O(\Delta\mathcal{H})^2\right) = \\
&= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} \left(1 + \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}}) \Delta\mathcal{H} + O(\Delta\mathcal{H})^2\right)
\end{aligned} \tag{I.2.2}$$

where $e^{\mathcal{H}_0} = Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}}, \boldsymbol{x}^{\text{s}} \hat{\boldsymbol{x}}^{\text{s}})$ and $\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} = Q_0(\boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$, therefore $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}}) = e^{\mathcal{H}_0} / \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}$. The reference distribution for the perturbative expansion $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$ is chosen as the LNA, i.e. a Gaussian conditional distribution centered around the deterministic steady states. We need bulk conditional probabilities, since, to define an effective action as such, we fix subnetwork variables and condition on them. The perturbative corrections arise from cubic terms in the action: let us consider them explicitly

$$\begin{aligned}
\Delta\mathcal{H} = \int_0^T dt \Bigg\{ \sum_i \mathrm{i}\hat{x}_i(t) \Big[\sum_{j,l,j\neq l} k^+_{ij,l} \delta x_i(t) \delta x_j(t) - \frac{1}{2} \sum_{j,l,j\neq l} k^+_{jl,i} \delta x_j(t) \delta x_l(t) + \\
\sum_l k^+_{ii,l} \delta x_i(t) \delta x_i(t) - \frac{1}{2} \sum_j k^+_{jj,i} \delta x_j(t) \delta x_j(t) \Big] \Bigg\}
\end{aligned} \tag{I.2.3}$$

The indices can belong to bulk or subnetwork, giving different combinations; we introduce a vectorial notation

$$\begin{aligned}
-\Delta\mathcal{H} = \int_0^T dt \Big\{ \mathrm{i}\hat{\boldsymbol{x}}^{\text{s}\,T} \Big[\boldsymbol{K}^{\text{s,ss}} (\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{s}}) + \boldsymbol{K}^{\text{s,sb}} (\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{b}}) + \boldsymbol{K}^{\text{s,bb}} (\delta\boldsymbol{x}^{\text{b}} \circ \delta\boldsymbol{x}^{\text{b}}) \Big] + \\
\mathrm{i}\hat{\boldsymbol{x}}^{\text{b}\,T} \Big[\boldsymbol{K}^{\text{b,ss}} (\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{s}}) + \boldsymbol{K}^{\text{b,sb}} (\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{b}}) + \boldsymbol{K}^{\text{b,bb}} (\delta\boldsymbol{x}^{\text{b}} \circ \delta\boldsymbol{x}^{\text{b}}) \Big] \Big\}
\end{aligned} \tag{I.2.4}$$

making use of the Hadamard product $\circ$ and $\boldsymbol{K}^{\text{s,sb}}$, $\boldsymbol{K}^{\text{b,sb}}$, $\boldsymbol{K}^{\text{b,bb}}$, $\boldsymbol{K}^{\text{b,ss}}$, $\boldsymbol{K}^{\text{s,bb}}$ as shorthands for the nonlinear couplings. As explained in the main text, we do not consider all the possible reactions occurring between subnetwork and bulk, thus we set $\boldsymbol{K}^{\text{b,ss}} = \boldsymbol{K}^{\text{s,bb}} \equiv 0$. We have also already dropped δ from $\hat{\boldsymbol{x}}^{\text{s}}$ and $\hat{\boldsymbol{x}}^{\text{b}}$ as the marginal auxiliary mean $\hat{\boldsymbol{\mu}}^{\text{s}} = \hat{\boldsymbol{\mu}}^{\text{b}} \equiv 0$.

The first order correction (as clear from the second term in (I.2.2)) reduces to an average w.r.t. the conditional distribution $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$. This average is equivalent to substituting combinations

of bulk variables in (I.2.3) by conditional Gaussian moments which, by Wick's theorem, can be expressed in terms of means and correlations.

$$-\langle\Delta\mathcal{H}\rangle = \int_0^T dt\Big\{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}\Big[\boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{x}^{\mathrm{s}}) + \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}})\Big] + \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}\,T}\Big[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}) + \boldsymbol{K}^{\mathrm{b,bb}}(\delta\boldsymbol{\mu}^{\mathrm{b|s}}\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}} + \boldsymbol{C}^{\mathrm{bb|s}}(t,t))\Big]\Big\} \tag{I.2.5}$$

where one uses that the conditional average $\langle\delta\boldsymbol{x}^{\mathrm{b}}(t)\delta\boldsymbol{x}^{\mathrm{b}\,T}(t)\rangle = \delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\delta\boldsymbol{\mu}^{\mathrm{b|s}\,T}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t)$, with $\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\mu}^{\mathrm{b|s}}(t) - \boldsymbol{\mu}^{\mathrm{b}}$. In developing the Wick's theorem for moments such as $\sim \langle\hat{\boldsymbol{x}}^{\mathrm{b}}\delta\boldsymbol{x}^{\mathrm{b}}\delta\boldsymbol{x}^{\mathrm{b}}\rangle$ or $\sim \langle\hat{\boldsymbol{x}}^{\mathrm{b}}\delta\boldsymbol{x}^{\mathrm{b}}\delta\boldsymbol{x}^{\mathrm{s}}\rangle$ we have used that the conditional equal time response functions $\boldsymbol{R}^{\mathrm{bb|s}}(t,t) = 0$ vanish. We next substitute the expression for $\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}$ (I.1.10)

$$-\langle\Delta\mathcal{H}\rangle = \int_0^T dt\Big\{\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\Big[\boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t)) + \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t))\Big] + \int_t^T dt'\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t')\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t)}\Big[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)) + \boldsymbol{K}^{\mathrm{b,bb}}(\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t))\Big]\Big\} \tag{I.2.6}$$

and we re-instate the time dependence as important to define the memory. First, we note that $\boldsymbol{K}^{\mathrm{s,ss}}$ appearing in the first line of (I.2.6) gives the nonlinear subnetwork dynamics, which is local in time. The reduced dynamics for $\boldsymbol{x}^{\mathrm{s}}(t)$ includes a memory $\boldsymbol{M}^{\mathrm{ss}}(t,t')$ and a coloured noise $\boldsymbol{\chi}(t)$ with covariance

$$\langle\boldsymbol{\chi}(t)\boldsymbol{\chi}^{T}(t')\rangle = \boldsymbol{N}^{\mathrm{ss}}(t,t') \tag{I.2.7}$$

to be determined. One can write $\boldsymbol{N}^{\mathrm{ss}}(t,t') = \boldsymbol{N}_0^{\mathrm{ss}}(t,t') + \boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ as the sum of a purely Gaussian term and a first order correction. Similarly the memory comprises two terms: $\boldsymbol{M}^{\mathrm{ss}}(t-t')$, which like $\boldsymbol{N}_0^{\mathrm{ss}}(t,t')$ can be calculated starting from the quadratic part of the action (as shown in Appendix H) and $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')$, which like $\boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ contains the contributions of cubic terms treated by means of perturbation theory. To evaluate these first order corrections, we insert in (I.2.6) the expressions for $\boldsymbol{\mu}^{\mathrm{b|s}}(t)$ (5.5.9): to illustrate the procedure we focus on the first term in the second line of (I.2.6), which becomes

$$\int_0^T dt'\int_{t'}^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t')) = \int_0^T dt'\int_{t'}^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\Big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\int_0^{t'}dt''e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'') - \delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\int_0^T dt''\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^{T}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'')\Big) \tag{I.2.8}$$

The first of the resulting two terms contributes to the reduced subnetwork dynamics via a temporal integral defining a nonlinear memory term given by

$$\int_0^t dt'\int_0^{t'}dt''\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')) \tag{I.2.9}$$

This describes how *products* of subnetwork concentrations feed back into the evolution of a single concentration. More generally, one can see that terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$ define the nonlinear memory function as given by (5.5.13) of the main text (where the kernel of (I.2.9) appears in the second line). In the second term of (I.2.8), the factor multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)$ and $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'')$

$$\int_0^t dt'\,\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T) \tag{I.2.10}$$

contributes to the covariance of the effective noise (which always enters the MSRJD formalism via quadratic terms in the auxiliary variables, as is clear from (5.2.12)) with an additional $\boldsymbol{x}^{\mathrm{s}}$-dependence. Treating the other terms in (I.2.6) in the same way and dropping third order terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$ that encode non-vanishing higher cumulants of the noise distribution one obtains the perturbative contribution to the effective noise covariance all terms of the form $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$; explicitly

$$\begin{aligned}\boldsymbol{N}_1^{\mathrm{ss}}(t,t'') &= \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{C}^{\mathrm{bb|s}}(t,t'')(\boldsymbol{K}^{\mathrm{sb}})^T)+\int_0^t dt'\,\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T)\\ &+ \int_0^t ds\int_0^s dt'\,\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-s)}\Big[\boldsymbol{K}^{\mathrm{b,bb}}(e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T)+\\ &\qquad \boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T\circ e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'))\Big]\end{aligned} \tag{I.2.11}$$

For the sake of completeness, one should also include corrections stemming from the fact that the white noise covariance $\boldsymbol{\Sigma}(\boldsymbol{x})$ as given by (5.2.4) does *depend* on concentrations. Since in this work we do not discuss further those terms, we prefer to analyze them in a forthcoming paper, where the full investigation of stochastic effects will be provided. Finally terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$ indicate noise non-Gaussian features we are not interested in for the purpose of the GVA.

Appendix J

Projection methods vs GVA

J.1 Projection methods vs GVA

The application of projection methods to protein-protein interaction networks has been studied by Rubin et al. in [77]: here we summarize some basic aspect necessary to develop a comparison with the GVA.

One is interested in deriving a closed-form expression for the memory function and the random force both for the linearized and the fully nonlinear dynamics: in the second case, it is necessary to appeal to the limit of vanishing noise. As the variance of copy number fluctuations scales as $\epsilon = 1/V$, i.e. the inverse of the reaction volume, the contribution of noise can be considered negligible (corresponding to $\epsilon \to 0$) for suitably large reaction volumes (see section 5.2). While one needs a nonzero ϵ for initially applying the Zwanzig-Mori formalism, the authors always then take the small ϵ limit. Let us stress that the small ϵ limit is not necessary for the linear dynamics, as the noise drops out from the equations for conditionally averaged concentrations whatever the value of ϵ.

In addition, to evaluate the memory and the random force from projection operators, one needs the steady state distribution for the deviations $\delta\boldsymbol{x}$. If the noise is negligible ($\epsilon \to 0$), $\delta\boldsymbol{x}$ is small and one can find their steady state distribution by linearizing the Langevin equation around $\delta\boldsymbol{x} = 0$; as a result, the steady state distribution of $\delta\boldsymbol{x}$ is a Gaussian and its covariance satisfies the Lyapunov equation. The structure of the covariance is not unique, but fixed by the type of fluctuations. The choice of independent Poisson fluctuations for each species, thus a covariance with a diagonal structure, produces the simplest projected equations. These describe the evolution for the subnetwork conditionally on the available knowledge of the initial conditions which are specified, for the bulk, via some probability distribution. The solution is the mean trajectory w.r.t. to this initial

distribution, from which thus stochastic fluctuations are averaged out: we will use for it the same notation as for the single instances $\boldsymbol{x}^{\mathrm{s}}(t)$ as they coincide in the limit $\epsilon \to 0$. As a consequence, we expect that the equations of motion given by projection methods for such conditionally averaged concentrations agree with the noise averaged subnetwork equations in the GVA: let us look closely at this comparison.

J.2 Linearized projected equations

The starting point is the Langevin dynamics (5.2.1) and its corresponding Fokker-Planck equation, encoded by $\mathcal{L}$, the so-called adjoint Fokker-Planck operator. Referring to [77] for the entire derivation, we shall directly provide the final form of the linearized dynamics

$$\frac{d\delta x_i(t)}{dt} = \sum_{j=1}^{N^{\mathrm{s}}} \delta x_j(t) \boldsymbol{\Omega}^{\mathrm{ss}}_{ji} + \sum_{j=1}^{N^{\mathrm{s}}} \int_0^t dt' \delta x_j(t') \boldsymbol{M}^{\mathrm{ss}}_{ji}(t-t') + r_i(t) \tag{J.2.1}$$

This is obtained by applying two operators projecting either onto the subspace of subnetwork d.o.f. or onto the orthogonal one; in the linearized dynamics, it is found that they can be represented by matrices with a simple block structure, $\boldsymbol{P}$ and $\boldsymbol{Q}$ respectively

$$\boldsymbol{P} = \begin{pmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{pmatrix}$$

$$\boldsymbol{Q} = \begin{pmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{pmatrix}$$

The adjoint Fokker-Planck operator $\mathcal{L}$ can be cast in matrix form as well denoted as $\boldsymbol{L}$. Note that $\boldsymbol{L} = \boldsymbol{K}^T$, i.e. it is equivalent to the transpose of dynamical matrix of the GVA; similarly, it can be partitioned into blocks for the bulk and the subnetwork part

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{\mathrm{ss}} & \boldsymbol{L}^{\mathrm{sb}} \\ \boldsymbol{L}^{\mathrm{bs}} & \boldsymbol{L}^{\mathrm{bb}} \end{pmatrix}$$

Exploiting this correspondence between operators and matrices one has for the linear dynamics

- $\boldsymbol{\Omega}$ is the top left block (related to subnetwork variables) of $\boldsymbol{PL}$, thus $\boldsymbol{\Omega} = \boldsymbol{L}^{\mathrm{ss}}$.
- $\boldsymbol{M}(t-t')$ is the top left block of $\boldsymbol{PLQ}e^{\boldsymbol{QLQ}(t-t')}\boldsymbol{QL}$, i.e.

$$\boldsymbol{M}(t-t') = \boldsymbol{L}^{\mathrm{sb}} e^{\boldsymbol{L}^{\mathrm{bb}}(t-t')} \boldsymbol{L}^{\mathrm{bs}} \tag{J.2.2}$$

- The random force is given by the s entries of $\delta \boldsymbol{x}^{\mathrm{b}\,T}(0) e^{\boldsymbol{QLQ}t}\boldsymbol{QL}$, i.e.

$$\boldsymbol{r}_0^T(t) = \delta \boldsymbol{x}^{\mathrm{b}\,T}(0) e^{\boldsymbol{L}^{\mathrm{bb}}t} \boldsymbol{L}^{\mathrm{bs}} \tag{J.2.3}$$

J.3 Nonlinear projected equations

Also in this case, as a starting point one resorts to suitable matrix representations for the operators involved. Even when a subnetwork has been selected, still there is arbitrariness about the space onto which project. Rubin et al. show [77] that the best option is to project onto protein concentrations and products of concentrations.

The nonlinearity is represented as a linear coupling between a concentration and products of other two; the projected equations have the form

$$\begin{aligned}\frac{d\delta x_i(t)}{dt} = &\sum_{j=1}^{N^s} \delta x_j(t)\boldsymbol{\Omega}^{\mathrm{ss}}_{ji} + \sum_{1\le j\le k\le N^s} \delta x_j(t)\delta x_k(t)\boldsymbol{\Omega}^{\mathrm{ss,s}}_{(jk),i} + \int_0^t dt'\Big(\sum_{j=1}^{N^s} \delta x_j(t')\boldsymbol{M}^{\mathrm{ss}}_{ji}(t-t')+ \\ &+ \sum_{1\le j\le k\le N^s} \delta x_j(t')\delta x_k(t')\boldsymbol{M}^{\mathrm{ss,s}}_{(jk),i}(t-t')\Big) + r_i(t)\end{aligned} \tag{J.3.1}$$

Focussing on an observable z_α (summarizing both simple concentrations and products), one can write the adjoint Fokker-Planck operator $\mathcal{L}$ in a matrix form such that

$$\partial_t z_\alpha = \sum_\beta z_\beta L_{\beta\alpha} + \delta x^3 + O(\epsilon) \tag{J.3.2}$$

δx^3 represents cubic terms, which are not captured at this order of accuracy, while $O(\epsilon)$ vanish in the small ϵ limit. The matrix representation therefore mirrors this choice of an enlarged space containing also products and it is valid for small ϵ. It reads explicitly as follows

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{\mathrm{ss}} & \boldsymbol{L}^{\mathrm{sb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{bs}} & \boldsymbol{L}^{\mathrm{bb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{ss,s}} & \boldsymbol{L}^{\mathrm{ss,b}} & \boldsymbol{L}^{\mathrm{ss,ss}} & \boldsymbol{L}^{\mathrm{ss,sb}} & \boldsymbol{L}^{\mathrm{ss,bb}} \\ \boldsymbol{L}^{\mathrm{sb,s}} & \boldsymbol{L}^{\mathrm{sb,b}} & \boldsymbol{L}^{\mathrm{sb,ss}} & \boldsymbol{L}^{\mathrm{sb,sb}} & \boldsymbol{L}^{\mathrm{sb,bb}} \\ \boldsymbol{L}^{\mathrm{bb,s}} & \boldsymbol{L}^{\mathrm{bb,b}} & \boldsymbol{L}^{\mathrm{bb,ss}} & \boldsymbol{L}^{\mathrm{bb,sb}} & \boldsymbol{L}^{\mathrm{bb,bb}} \end{pmatrix}$$

$\boldsymbol{L}$ consists of 5 rows and columns, referring to linear subnetwork concentrations (s), subnetwork products (ss), mixed subnetwork-bulk products (sb), to be considered as bulk elements, and products of bulk concentrations (bb). The dynamics for the products, contained in the bottom right blocks ($\boldsymbol{L}^{\mathrm{ss,ss}}$, $\boldsymbol{L}^{\mathrm{ss,sb}}$ etc.), is simply derived from linearized dynamics, thus

$$\partial_t(\delta x_i \delta x_j) = \delta x_j \partial_t \delta x_i + \delta x_i \partial_t \delta x_j \tag{J.3.3}$$

In other words, in the evolution of a product only products appear: this explains why the top right block vanishes, i.e. applying $\boldsymbol{L}$ to quadratic observables does not give linear terms. All the coefficients in the bottom right blocks are simply brought forward from the linearized dynamics.

Nonlinearities enter instead via the bottom left blocks ($\boldsymbol{L}^{\mathrm{ss,s}}$, $\boldsymbol{L}^{\mathrm{ss,b}}$ etc.), containing the coefficients multiplying products in the equations for linear observables. For projection matrices one has

$$\boldsymbol{P} = \begin{pmatrix} \mathbb{1} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{1} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \end{pmatrix}$$

while $\boldsymbol{Q}$ is analogous to $\boldsymbol{P}$, but the roles of $\mathbb{1}$ and $\mathbb{0}$ along the diagonal are interchanged

$$\boldsymbol{Q} = \begin{pmatrix} \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{1} \end{pmatrix}$$

The combinations of matrices one needs, to generalize the formulas of the previous section, are $\boldsymbol{PL}$, $\boldsymbol{QLQ}$ and $\boldsymbol{QL}$, namely,

$$\boldsymbol{PL} = \begin{pmatrix} \boldsymbol{L}^{\mathrm{ss}} & \boldsymbol{L}^{\mathrm{sb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{ss,s}} & \boldsymbol{L}^{\mathrm{ss,b}} & \boldsymbol{L}^{\mathrm{ss,ss}} & \boldsymbol{L}^{\mathrm{ss,sb}} & \boldsymbol{L}^{\mathrm{ss,bb}} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \end{pmatrix}$$

$$\boldsymbol{QLQ} = \begin{pmatrix} \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \boldsymbol{L}^{\mathrm{bb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \boldsymbol{L}^{\mathrm{sb,b}} & \mathbb{0} & \boldsymbol{L}^{\mathrm{sb,sb}} & \boldsymbol{L}^{\mathrm{sb,bb}} \\ \mathbb{0} & \boldsymbol{L}^{\mathrm{bb,b}} & \mathbb{0} & \boldsymbol{L}^{\mathrm{bb,sb}} & \boldsymbol{L}^{\mathrm{bb,bb}} \end{pmatrix}$$

$$\boldsymbol{QL} = \begin{pmatrix} \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{bs}} & \boldsymbol{L}^{\mathrm{bb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{sb,s}} & \boldsymbol{L}^{\mathrm{sb,b}} & \boldsymbol{L}^{\mathrm{sb,ss}} & \boldsymbol{L}^{\mathrm{sb,sb}} & \boldsymbol{L}^{\mathrm{sb,bb}} \\ \boldsymbol{L}^{\mathrm{bb,s}} & \boldsymbol{L}^{\mathrm{bb,b}} & \boldsymbol{L}^{\mathrm{bb,ss}} & \boldsymbol{L}^{\mathrm{bb,sb}} & \boldsymbol{L}^{\mathrm{bb,bb}} \end{pmatrix}$$

We emphasize that $\boldsymbol{QLQ}$ has lower triangular block structure, thus $e^{\boldsymbol{QLQ}t}$ has the same structure with diagonal blocks which are the exponentials of those in $\boldsymbol{QLQ}$, i.e. for example $[e^{\boldsymbol{QLQ}t}]_{\mathrm{bb}} =$

$e^{\boldsymbol{L}^{\mathrm{bb}}t}$. Another property that can be deduced from such a structure is

$$\boldsymbol{E}_{\mathrm{sb,b}}(t) = \int_0^t dt' \boldsymbol{E}_{\mathrm{sb,sb}}(t-t')\boldsymbol{L}^{\mathrm{sb,b}}\boldsymbol{E}_{\mathrm{bb}}(t') \tag{J.3.4a}$$

$$\boldsymbol{E}_{\mathrm{bb,b}}(t) = \int_0^t dt' \boldsymbol{E}_{\mathrm{bb,bb}}(t-t')\boldsymbol{L}^{\mathrm{bb,b}}\boldsymbol{E}_{\mathrm{bb}}(t') \tag{J.3.4b}$$

where we have introduced the notation $e^{\boldsymbol{QLQ}(t-t')} = \boldsymbol{E}(t-t')$.

The rate constants, the memory function and the random force for the full nonlinear dynamics can be found in analogy with linear case provided that we consider these enlarged matrices; therefore

- The nonlinear rate matrix for the internal subnetwork dynamics is $\boldsymbol{\Omega}^{\mathrm{ss,s}} = \boldsymbol{L}^{\mathrm{ss,s}}$.
- The memory is the ss block of $\boldsymbol{PLQ}e^{\boldsymbol{QLQ}t}\boldsymbol{QL}$, whose nonlinear part is

  $$\boldsymbol{M}^{\mathrm{ss,s}}(t-t') = \boldsymbol{L}^{\mathrm{ss,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\mathrm{\cdot b,\cdot b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\mathrm{\cdot b,\cdot b}}(t-t')\boldsymbol{L}^{\mathrm{\cdot b,s}} \tag{J.3.5}$$

  where "·b" is meant as joining "sb" and "bb" ranges. Assumptions about what reactions between the bulk and the subnetwork can occur imply that some $\boldsymbol{L}$ blocks are zero, namely: $\boldsymbol{L}^{\mathrm{ss,bb}} = \boldsymbol{L}^{\mathrm{bb,ss}} = \boldsymbol{L}^{\mathrm{ss,b}} = \boldsymbol{L}^{\mathrm{bb,s}} = \mathbb{0}$. These constraints then simplify the expression for memory considerably

  $$\boldsymbol{M}^{\mathrm{ss,s}}(t-t') = \boldsymbol{L}^{\mathrm{ss,sb}}\boldsymbol{E}_{\mathrm{sb,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,sb}}\boldsymbol{E}_{\mathrm{sb,sb}}(t-t')\boldsymbol{L}^{\mathrm{sb,s}} \tag{J.3.6}$$

  We shall in any case stick to the most general case. The memory function in the dynamics for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ is then embedded in a time integral over the past history of all possible subnetwork products $\delta\boldsymbol{x}^{\mathrm{ss}}(t')$, thus the general nonlinear memory for the projected equations is

  $$\boldsymbol{\mathcal{M}}^T_{\mathrm{proj}}(t) = m_1 + m_2 + m_3 \tag{J.3.7}$$

  with

  $$m_1 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} \tag{J.3.8}$$

  $$m_2 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\mathrm{\cdot b,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} \tag{J.3.9}$$

  $$m_3 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\mathrm{\cdot b,\cdot b}}(t-t')\boldsymbol{L}^{\mathrm{\cdot b,s}} \tag{J.3.10}$$

- Rigorously the random force cannot be calculated in closed form from the matrix representation introduced, but a vectorial expression can be still provided up to the quadratic contribution: higher order terms, not captured by the latter, are expected anyway to be small or negligible. We shall write $\boldsymbol{r}(t) = \boldsymbol{r}_0(t) + \boldsymbol{r}_1(t)$, $\boldsymbol{r}_1(t)$ being the nonlinear random force

  $$\boldsymbol{r}_1^T(t) = \delta\boldsymbol{x}^{\mathrm{sb}\,T}(0)\left[\boldsymbol{E}_{\mathrm{sb,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\mathrm{sb,sb}}(t)\boldsymbol{L}^{\mathrm{sb,s}}\right] + \delta\boldsymbol{x}^{\mathrm{bb}\,T}(0)\left[\boldsymbol{E}_{\mathrm{bb,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\mathrm{bb,sb}}(t)\boldsymbol{L}^{\mathrm{sb,s}}\right] \tag{J.3.11}$$

From these results, we see that the projection method divides the non-Markovian contribution from the bulk into the random force, depending only on initial conditions, and the memory, containing all the single points in the past for single species and products. This structural feature is important for the comparison to the GVA, which does not separate so neatly the non-local-in-time terms but still gives an *equivalent* approximation *up to the second order and in the limit* $\epsilon \to 0$. To provide the tools for this comparison, we first recall the perturbative expansion allowing us to derive nonlinear corrections in the Gaussian variational approach (section J.4) and we explicitly derive this equivalence in section J.5.

J.4 Full nonlinear memory and random force in the GVA

As we mentioned, the GVA and the projection method are expected to agree only for $\epsilon \to 0$: we then restrict ourselves to this case, i.e. the deterministic dynamics. We imagine translating the perturbative approach of section I.2 in appendix I to a notation analogous to the one in the previous section J.3; in particular the cubic part of the action (I.2.4) would read

$$-\Delta\mathcal{H} = \int_0^T dt\Big\{\Big[(\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{x}^{\mathrm{s}})^T\boldsymbol{L}^{\mathrm{ss,s}} + (\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{x}^{\mathrm{b}})^T\boldsymbol{L}^{\mathrm{sb,s}} + (\delta\boldsymbol{x}^{\mathrm{b}}\circ\delta\boldsymbol{x}^{\mathrm{b}})^T\boldsymbol{L}^{\mathrm{bb,s}}\Big]\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}} + \Big[(\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{x}^{\mathrm{s}})^T\boldsymbol{L}^{\mathrm{ss,b}} + (\delta\boldsymbol{x}^{\mathrm{s}}\circ\delta\boldsymbol{x}^{\mathrm{b}})^T\boldsymbol{L}^{\mathrm{sb,b}} + (\delta\boldsymbol{x}^{\mathrm{b}}\circ\delta\boldsymbol{x}^{\mathrm{b}})\boldsymbol{L}^{\mathrm{bb,b}}\Big]\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}\Big\} \tag{J.4.1}$$

In light of the fact that we allow only certain processes, in (I.2.4) we had invoked the simplification $\boldsymbol{L}^{\mathrm{ss,b}} = \boldsymbol{L}^{\mathrm{bb,s}} \equiv 0$, while here, for the sake of a rigorous comparison, we keep all the nonlinear couplings to make the argument as general as possible. The nonlinear memory and effective noise covariance are evaluated by taking the average w.r.t. the Gaussian conditional distribution over bulk variables, as given by (I.2.6). In the new notation it becomes

$$-\langle\Delta\mathcal{H}\rangle = \int_0^T dt(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t))^T\boldsymbol{L}^{\mathrm{ss,s}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \tag{J.4.2}$$
$$\int_0^T dt\Big\{\Big[(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t))^T\boldsymbol{L}^{\mathrm{sb,s}} + (\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t))^T\boldsymbol{L}^{\mathrm{bb,s}}\Big]\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) +$$
$$\int_t^T dt'\Big[(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t))^T\boldsymbol{L}^{\mathrm{sb,b}} + (\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t))^T\boldsymbol{L}^{\mathrm{bb,b}}\Big]\boldsymbol{L}^{\mathrm{bs}}\boldsymbol{E}_{\mathrm{bb}}(t'-t)\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t')\Big\}$$

given that the auxiliary conditional mean, from (5.5.7), can be rewritten as

$$\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) = \int_t^T dt'\,\boldsymbol{E}_{\mathrm{bb}}(t'-t)\boldsymbol{L}^{\mathrm{bs}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{J.4.3}$$

Then one substitutes the expressions for the conditional means (5.5.9) - in this notation

$$\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\mu}^{\mathrm{b|s}}(t) - \boldsymbol{\mu}^{\mathrm{b}}(t) = \int_0^t dt'\,e^{(\boldsymbol{L}^{\mathrm{bb}})^T(t-t')}(\boldsymbol{L}^{\mathrm{sb}})^T\delta\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt'\boldsymbol{C}^{\mathrm{bb|s}}(t,t')\boldsymbol{L}^{\mathrm{bs}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{J.4.4}$$

in (J.4.2) and by isolating the terms multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)$ one can read the nonlinear reduced dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$. The memory contribution of the GVA comes from the terms quadratic in $\delta\boldsymbol{x}^{\mathrm{s}}$ and non-local-in-time

$$\begin{aligned}\boldsymbol{\mathcal{M}}^{T}_{\mathrm{GVA}}(t) \;=\; & \int_0^t dt' \Big\{ (\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t-t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t))\boldsymbol{L}^{\mathrm{sb,s}} + \\ & \int_{t'}^t dt'' (\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t''-t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t''))\boldsymbol{L}^{\mathrm{sb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t'')\boldsymbol{L}^{\mathrm{bs}} + \\ & \int_0^t dt'' \int_{\max(t',t'')}^t ds\, (\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t'')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t''))\boldsymbol{L}^{\mathrm{bb,s}} + \\ & \int_0^t dt'' \int_{\max(t',t'')}^t ds\, (\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t'')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t''))\boldsymbol{L}^{\mathrm{bb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-s)\boldsymbol{L}^{\mathrm{bs}} \Big\}\end{aligned} \tag{J.4.5}$$

This corresponds to the double time integral of (5.5.13) in the modified notation and keeping all the couplings. Similarly from the terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$ we can deduce the effective noise covariance

$$\begin{aligned}\boldsymbol{N}_1^{\mathrm{ss}}(t,t'') \;=\; & ((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t))\boldsymbol{L}^{\mathrm{sb,s}} + \\ & \int_0^t dt' ((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t',t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t'))\boldsymbol{L}^{\mathrm{sb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \\ & + \int_0^t dt' \Big[(\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t-t') \circ (\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t,t''))\boldsymbol{L}^{\mathrm{bb,s}} + \\ & \qquad ((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t-t'))\boldsymbol{L}^{\mathrm{bb,s}} \Big] \\ & + \int_0^t ds \int_0^s dt' \Big[(\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t') \circ (\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t',t''))\boldsymbol{L}^{\mathrm{bb,b}} + \\ & \qquad ((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t',t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t'))\boldsymbol{L}^{\mathrm{bb,b}} \Big] \boldsymbol{E}_{\mathrm{bb}}(t-s)\boldsymbol{L}^{\mathrm{bs}}\end{aligned} \tag{J.4.6}$$

an extension of (I.2.11). For $\epsilon \to 0$, the bulk conditional correlator is simply

$$\boldsymbol{C}^{\mathrm{bb|s}}(t',t'') = \boldsymbol{E}^{T}_{\mathrm{bb}}(t')\boldsymbol{C}^{\mathrm{bb}}(0,0)\boldsymbol{E}_{\mathrm{bb}}(t'') \tag{J.4.7}$$

as can be deduced from (5.5.11).

Let us stress that projection methods expand the random force, while our perturbative approach expands the correlator of a coloured noise. For the sake of comparison, it is therefore convenient to think in terms of the effective coloured noise with covariance $\boldsymbol{N}_0^{\mathrm{ss}} + \boldsymbol{N}_1^{\mathrm{ss}}$. $\boldsymbol{N}_1^{\mathrm{ss}}$ can be read from (J.4.6), while $\boldsymbol{N}_0^{\mathrm{ss}}$, the effective noise covariance of the linearized dynamics, is given by (5.4.7); we rewrite it as

$$\boldsymbol{N}_0^{\mathrm{ss}}(t,t') = (\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{E}^{T}_{\mathrm{bb}}(t)\boldsymbol{C}^{\mathrm{bb}}(0,0)\boldsymbol{E}_{\mathrm{bb}}(t')\boldsymbol{L}^{\mathrm{bs}} \tag{J.4.8}$$

Note that we have dropped the intrinsic noise contribution $\boldsymbol{\Sigma}^{\mathrm{ss}}$, $\boldsymbol{\Sigma}^{\mathrm{bb}}$ as we focus on $\epsilon \to 0$ as well as the explicit statement of the subnetwork initial condition $\boldsymbol{C}^{\mathrm{ss}}(0,0)\delta(t)\delta(t')$. The Gaussian noise $\boldsymbol{\chi}_0(t)$ such that $\langle \boldsymbol{\chi}_0(t)\boldsymbol{\chi}_0^T(t')\rangle = \boldsymbol{N}_0^{\mathrm{ss}}(t,t')$ is therefore

$$\boldsymbol{\chi}_0^T(t) = \delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t)\boldsymbol{L}^{\mathrm{bs}} \tag{J.4.9}$$

If we define $\boldsymbol{\chi}_1(t)$ as follows

$$\begin{aligned}\boldsymbol{\chi}_1^T(t) = &(\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t) \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t))\boldsymbol{L}^{\mathrm{sb,s}} + \\ &\int_0^t dt'\Big[(\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t'))\boldsymbol{L}^{\mathrm{sb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + (\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t) \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t-t'))\boldsymbol{L}^{\mathrm{bb,s}}\Big] \\ &+ \int_0^t ds \int_0^s dt'\Big[(\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(s-t'))\boldsymbol{L}^{\mathrm{bb,b}}\Big]\boldsymbol{E}_{\mathrm{bb}}(t-s)\boldsymbol{L}^{\mathrm{bs}}\end{aligned} \tag{J.4.10}$$

then the correlation function of $\boldsymbol{\chi}_0 + \boldsymbol{\chi}_1$ is given by $\boldsymbol{N}_0^{\mathrm{ss}} + \boldsymbol{N}_1^{\mathrm{ss}}$ at the linear order in $\delta\boldsymbol{x}^{\mathrm{s}}$ (this is the order in $\delta\boldsymbol{x}^{\mathrm{s}}$ to which $\boldsymbol{N}_1^{\mathrm{ss}}$ is calculated). For example, let us take the first term in (J.4.6): it is derived from cross-correlations of $\boldsymbol{\chi}_0$ (J.4.9) and the first term of $\boldsymbol{\chi}_1$ (J.4.10) as follows

$$\Big\langle((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{E}_{\mathrm{bb}}^T(t)\delta\boldsymbol{x}^{\mathrm{b}}(0)\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t''))\boldsymbol{L}^{\mathrm{sb,s}}\Big\rangle = ((\boldsymbol{L}^{\mathrm{bs}})^T\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') \circ \delta\boldsymbol{x}^{\mathrm{s}\,T}(t''))\boldsymbol{L}^{\mathrm{sb,s}} \tag{J.4.11}$$

by using the definition (J.4.7).

From expression (J.4.2) (and also from (I.2.5) although we did not comment about it), it can be seen that the Gaussian effective dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ exhibits also the $\boldsymbol{x}^{\mathrm{s}}$-independent term

$$\boldsymbol{C}^{\mathrm{bb|s}}(t,t)\boldsymbol{L}^{\mathrm{bb,s}} + \int_0^t dt'\boldsymbol{C}^{\mathrm{bb|s}}(t',t')\boldsymbol{L}^{\mathrm{bb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} = \langle\boldsymbol{\psi}_1^T(t)\rangle \tag{J.4.12}$$

which can be written as the conditional average of a vector $\boldsymbol{\psi}_1(t)$ defined as follows

$$\begin{aligned}\boldsymbol{\psi}_1^T(t) =&(\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t) \circ\ \delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t))\boldsymbol{L}^{\mathrm{bb,s}} + \\ &\int_0^t dt'(\delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t') \circ\ \delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t'))\boldsymbol{L}^{\mathrm{bb,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}}\end{aligned} \tag{J.4.13}$$

The conditional average of $\boldsymbol{\psi}_1$ produces in the action a cubic term, $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{b}}\delta\boldsymbol{x}^{\mathrm{b}}$, while its variance would appear in the action via a term of 6th order, thus it is not present at the order we have done the calculation. From (J.4.13) we see that $\boldsymbol{\psi}_1(t)$ is a temporally correlated term in the reduced subnetwork dynamics which depends quadratically on $\delta\boldsymbol{x}^{\mathrm{b}}(0)$; it can be regarded as a further contribution to the nonlinear "random force" of the GVA, as by definition the random force summarizes the uncertainty on the bulk initial conditions. As a result, the systematic perturbative expansion up to the cubic order in the action gives it as

$$\tilde{\boldsymbol{r}}_{1\,\mathrm{GVA}}(t) = \boldsymbol{\chi}_1(t) + \langle\boldsymbol{\psi}_1(t)\rangle \tag{J.4.14}$$

If we consider the fluctuating version of this quantity, namely

$$\boldsymbol{r}_{1\,\mathrm{GVA}}(t) = \boldsymbol{\chi}_1(t) + \boldsymbol{\psi}_1(t) \tag{J.4.15}$$

we can prove the equivalence with the projection formalism.

J.5 Proof of the equivalence

Let us assume that the initial deviations from the means are proportional to some factor δ, $\delta\boldsymbol{x}^{\mathrm{s}}(0) \sim \delta$ and $\delta\boldsymbol{x}^{\mathrm{b}}(0) \sim \delta$. The solution $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ can be expanded in powers of δ and we want to find out to what extent projection and GVA give the same results for this expansion in the limit $\epsilon \to 0$. We expect that the two descriptions are equivalent at $O(\delta^2)$ as in the GVA we keep cubic terms in the effective action while in projection methods we keep all quadratic observables.

J.5.1 Linear order in δ

To work out the solution of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ in this case, we need to include in its dynamics the expressions for memory and random force at $O(\delta)$. By re-writing the linear memory (5.4.6) of the GVA in the notation of section J.3 one has

$$\boldsymbol{M}^{\mathrm{ss}}(t-t') = \boldsymbol{L}^{\mathrm{sb}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} \tag{J.5.1}$$

where we have used $e^{\boldsymbol{L}^{\mathrm{bb}}t} = \boldsymbol{E}_{\mathrm{bb}}(t)$. The effective coloured noise at this order for the GVA is given by (J.4.9). Expressions (J.5.1) and (J.4.9) are the same as from projection methods (J.2.2) and (J.2.3), thus the resulting expansion of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ is identical to $O(\delta)$.

J.5.2 Quadratic order in δ

In this case, we need to include in the dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ the memory and the effective noise/random force at $O(\delta^2)$. This includes linear memory terms acting on the second order part of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$, which will be the same for projection and GVA. The remaining terms are the nonlinear memory and nonlinear effective noise/random force evaluated to order $O(\delta^2)$.

Expressions for nonlinear memory and random force from the two approaches (respectively (J.3.7) and (J.4.5), (J.3.11) and (J.4.15)) do not coincide if taken *separately*; we nevertheless expect that their *combination* is actually equivalent (because there can be different ways of writing expansions that are correct at the same order, namely $O(\delta^2)$).

It is worth though some preliminary comment, as a guide through our reasoning and formal manipulations. First of all, the memory function (J.4.5) differs from the one of projection methods (J.3.7): in particular, the two subnetwork species in quadratic terms are calculated at two times of the past instead of the same time, additional integrals over time also appear. One can think of making the intermediate times disappear, thus of matching the one-time and two-times structures

in the algebra, by expressing explicitly the dependence on subnetwork initial values. Next we observe that in the structure of the enlarged (i.e. including quadratic observables) $\boldsymbol{L}$, the non-linearity of interactions enters genuinely only in the bottom left off-diagonal block. The bottom right (diagonal) block, which represents the coupling between products, does not provide any additional information with respect to the linear dynamics: it describes how the products propagate under the linear dynamics. Note that this is precisely what the perturbative expansion implements, i.e. it describes products under the linear evolution. Therefore, we can drop the purely non-linear block and try to write also the products of subnetwork variables and conditional means of the perturbative expansion as functions of the initial conditions in the subnetwork via the exponential of a matrix consisting of a linear and a quadratic block.

We need to substitute in the nonlinear memory and random force the first order deterministic solution for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ and $\delta\boldsymbol{x}^{\mathrm{b}}(t)$ to evaluate them consistently to $O(\delta^2)$: let us perform the substitution in both approaches and compare the results. The projection memory is evaluated at $O(\delta^2)$ using this expression

$$\delta\boldsymbol{x}^{\cdot\cdot\,T}(t) = \delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t} \tag{J.5.2}$$

"$\cdot\cdot$" refers to the block obtained by joining "ss","sb" and "bb" ranges here and everywhere below, without further specification. Also in the GVA we have to write the products as in (J.5.2) but with the following caveat. The GVA is obtained via the inclusion of the dynamics of conditional *averages*, where the average is taken also over bulk initial conditions. This, importantly, implies that $\delta\boldsymbol{x}^{\mathrm{b}}(0)$ should not be treated as a fluctuating quantity as instead (J.5.2) does. More explicitly, we recall that the conditional bulk variables can be written as $\delta\boldsymbol{\mu}^{\mathrm{b|s}\,T}(t) = \boldsymbol{\nu}^T(t) + \hat{\boldsymbol{\nu}}^T(t)$, where only the deterministic part $\boldsymbol{\nu}(t)$ gives contributes to the memory

$$\boldsymbol{\nu}^T(t) = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\,\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{J.5.3}$$

(from expression (5.5.9)). On the other hand, the evolution of the products introduced by (J.5.2) accounts for a solution for bulk variables as follows

$$\delta\boldsymbol{x}^{\mathrm{b}\,T}(t) = \delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t) + \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\,\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{J.5.4}$$

with also deviations in the initial conditions $\delta\boldsymbol{x}^{\mathrm{b}}(0)$. The nonlinear memory for the variational approach (J.4.5) can be written as

$$\begin{aligned}
\boldsymbol{\mathcal{M}}^T_{\mathrm{GVA}}(t) &= \int_0^t dt' \left\langle \delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t'} \right\rangle_{\cdot\cdot} \boldsymbol{L}^{\cdot\cdot,\mathrm{b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \left\langle \delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t} \right\rangle_{\cdot\mathrm{b}} \boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}} = \\
&= \int_0^t dt' (\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t'})_{\cdot\cdot}\, \boldsymbol{L}^{\cdot\cdot,\mathrm{b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + (\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t})_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}} - m'_4 = \\
&= m'_1 + m'_2 + m'_3 - m'_4
\end{aligned} \tag{J.5.5}$$

We have

$$m_1' = \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\mathrm{ss}} \boldsymbol{L}^{\mathrm{ss,b}} \boldsymbol{E}_{\mathrm{bb}}(t-t') \boldsymbol{L}^{\mathrm{bs}} \tag{J.5.6}$$

$$m_2' = \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} \boldsymbol{L}^{\cdot \mathrm{b,b}} \boldsymbol{E}_{\mathrm{bb}}(t-t') \boldsymbol{L}^{\mathrm{bs}} \tag{J.5.7}$$

$$m_3' = (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t})_{\cdot \mathrm{b}} \boldsymbol{L}^{\cdot \mathrm{b,s}} \tag{J.5.8}$$

m_4' has minus sign as it is a correction term needed for the substitution in the second line of (J.5.5) to be valid (in other words, to compensate for the use of (J.5.4) instead of (J.5.3)); namely

$$\begin{aligned}
m_4' = {} & (\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}^{\mathrm{bb}}(t) \circ \delta \boldsymbol{x}^{\mathrm{s}\, T}(t)) \boldsymbol{L}^{\mathrm{sb,s}} + \int_0^t dt' (\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t') \circ \delta \boldsymbol{x}^{\mathrm{s}\, T}(t')) \boldsymbol{L}^{\mathrm{sb,b}} \boldsymbol{E}_{\mathrm{bb}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \\
& \int_0^t dt' \Big[(\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t) \circ \delta \boldsymbol{x}^{\mathrm{s}\, T}(t') \boldsymbol{L}^{\mathrm{sb}} \boldsymbol{E}_{\mathrm{bb}}(t-t')) \boldsymbol{L}^{\mathrm{bb,s}} + (\delta \boldsymbol{x}^{\mathrm{s}\, T}(t') \boldsymbol{L}^{\mathrm{sb}} \boldsymbol{E}_{\mathrm{bb}}(t-t') \circ \delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t)) \boldsymbol{L}^{\mathrm{bb,s}} \Big] \\
& + (\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t) \circ \delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t)) \boldsymbol{L}^{\mathrm{bb,s}} + \int_0^t ds\, (\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(s) \circ \delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(s)) \boldsymbol{L}^{\mathrm{bb,b}} + \\
& \int_0^t ds \int_0^s dt' \Big[(\delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t') \circ \delta \boldsymbol{x}^{\mathrm{s}\, T}(t') \boldsymbol{L}^{\mathrm{sb}} \boldsymbol{E}_{\mathrm{bb}}(s-t')) \boldsymbol{L}^{\mathrm{bb,b}} + \\
& \qquad (\delta \boldsymbol{x}^{\mathrm{s}\, T}(t') \boldsymbol{L}^{\mathrm{sb}} \boldsymbol{E}_{\mathrm{bb}}(s-t') \circ \delta \boldsymbol{x}^{\mathrm{b}\, T}(0) \boldsymbol{E}_{\mathrm{bb}}(t')) \boldsymbol{L}^{\mathrm{bb,b}} \Big] \boldsymbol{E}_{\mathrm{bb}}(t-s) \boldsymbol{L}^{\mathrm{bs}}
\end{aligned} \tag{J.5.9}$$

Recalling equations (J.4.10), (J.4.13) and (J.4.15), one can write

$$m_4' = \boldsymbol{r}_{1\,\mathrm{GVA}}^T(t) \tag{J.5.10}$$

From (J.5.5) and (J.5.10) we have

$$m_1' + m_2' + m_3' = \boldsymbol{\mathcal{M}}_{\mathrm{GVA}}^T(t) + \boldsymbol{r}_{1\,\mathrm{GVA}}^T(t) \tag{J.5.11}$$

Let us now compare term by term the nonlinear memory from projection methods (J.3.7) and from the GVA (J.5.5): one has immediately $m_1 = m_1'$ (see (J.3.8) and (J.5.6)). We apply next the identity

$$(\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\mathrm{ss}} \boldsymbol{L}^{\mathrm{ss},\cdot \mathrm{b}} = \partial_{t'} (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} - (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} \boldsymbol{L}^{\cdot \mathrm{b},\cdot \mathrm{b}} \tag{J.5.12}$$

to manipulate the sum of m_2 (J.3.9) and m_3 (J.3.10), as follows

$$\begin{aligned}
m_2 + m_3 = {} & \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\mathrm{ss}} \boldsymbol{L}^{\mathrm{ss},\cdot \mathrm{b}} [\boldsymbol{E}_{\cdot \mathrm{b,b}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot \mathrm{b},\cdot \mathrm{b}}(t-t') \boldsymbol{L}^{\cdot \mathrm{b,s}}] = \\
= {} & \int_0^t dt' [\partial_{t'} (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} - (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} \boldsymbol{L}^{\cdot \mathrm{b},\cdot \mathrm{b}}] [\boldsymbol{E}_{\cdot \mathrm{b,b}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot \mathrm{b},\cdot \mathrm{b}}(t-t') \boldsymbol{L}^{\cdot \mathrm{b,s}}] = \\
= {} & - \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} \partial_{t'} [\boldsymbol{E}_{\cdot \mathrm{b,b}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot \mathrm{b},\cdot \mathrm{b}}(t-t') \boldsymbol{L}^{\cdot \mathrm{b,s}}] \\
& + (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} [\boldsymbol{E}_{\cdot \mathrm{b,b}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot \mathrm{b},\cdot \mathrm{b}}(t-t') \boldsymbol{L}^{\cdot \mathrm{b,s}}] \Big|_0^t \\
& - \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\, T}(0) e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot} t'})_{\cdot \mathrm{b}} \boldsymbol{L}^{\cdot \mathrm{b},\cdot \mathrm{b}} [\boldsymbol{E}_{\cdot \mathrm{b,b}}(t-t') \boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot \mathrm{b},\cdot \mathrm{b}}(t-t') \boldsymbol{L}^{\cdot \mathrm{b,s}}]
\end{aligned} \tag{J.5.13}$$

In the third line, an integration by parts has been implemented and $-\partial_{t'}$ can be further substituted by ∂_t as it acts on the difference $(t-t')$. Let us the consider the boundary term from this integration

$$\begin{aligned}(\delta\boldsymbol{x}^{\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot,\cdot}t})_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}} - \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}) \\ = m_3' - \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}})\end{aligned} \tag{J.5.14}$$

where we used (J.5.8), $\boldsymbol{E}_{\mathrm{sb},\mathrm{b}}(0) = \boldsymbol{E}_{\mathrm{bb},\mathrm{b}}(0) = \mathbb{0}$, $\boldsymbol{E}_{\mathrm{sb},\mathrm{sb}}(0) = \boldsymbol{E}_{\mathrm{bb},\mathrm{bb}}(0) = \mathbb{1}$ and that all the terms in $e^{\boldsymbol{L}^{\cdot,\cdot}t'}$ at $t' = 0$ give $\mathbb{1}$. Equation (J.5.13) can be rewritten

$$\begin{aligned}m_2 + m_3 = m_3' &- \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}) \\ &+ \int dt'(\delta\boldsymbol{x}^{\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot,\cdot}t'})_{\cdot\mathrm{b}}\partial_t[\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t')\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}] \\ &- \int dt'(\delta\boldsymbol{x}^{\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot,\cdot}t'})_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}[\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t')\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}]\end{aligned} \tag{J.5.15}$$

In the second line of (J.5.15), we can apply the identity

$$\partial_t\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t-t') = \boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t-t') + \boldsymbol{L}^{\cdot\mathrm{b},\mathrm{b}}\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{J.5.16}$$

which can be deduced from the properties (J.3.4) and from

$$\partial_t\boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t') = \boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}\boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t') \tag{J.5.17}$$

As a consequence, also by recalling (J.5.7), we obtain

$$m_2 + m_3 = m_3' - \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}) + m_2' \tag{J.5.18}$$

In light of (J.3.7) and (J.5.11), we have that, symbolically,

$$\boldsymbol{\mathcal{M}}^T_{\mathrm{GVA}}(t) + \boldsymbol{r}^T_{1\,\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}^T_{\mathrm{proj}}(t) + \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b},\mathrm{b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b},\mathrm{s}}) \tag{J.5.19}$$

the second term is exactly the nonlinear part of the transposed projection random force (J.3.11), which we here denote $\boldsymbol{r}^T_{1\,\mathrm{proj}}(t)$. Finally one obtains the equivalence at $O(\delta^2)$ we aimed to prove, i.e.

$$\boldsymbol{\mathcal{M}}_{\mathrm{GVA}}(t) + \boldsymbol{r}_{1\,\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}_{\mathrm{proj}}(t) + \boldsymbol{r}_{1\,\mathrm{proj}}(t) \tag{J.5.20}$$

The comparison becomes even simpler if the bulk is assumed initially at steady state, i.e. $\delta\boldsymbol{x}^{\mathrm{b}}(0) \equiv 0$, as all the random force terms would vanish and $\boldsymbol{\mathcal{M}}_{\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}_{\mathrm{proj}}(t)$.

Appendix K

Effective equations solver

K.1 Effective equations solver

Integro-differential equations such as (5.5.12) can be solved via any standard *differential* equations solver by appeal to an enlarged system of differential equations. The idea behind it is that every memory integral term can be seen as the solution of a differential equation. Let us start from the subnetwork reduced dynamics in the form

$$\frac{d}{dt}\delta\boldsymbol{x}^{\mathrm{s}}(t) = \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t)) + \mathcal{M}(t) \tag{K.1.1}$$

where we have taken the large volume limit, in order to neglect the intrinsic noise. By reconstructing the derivation of the memory vector $\mathcal{M}(t)$ in terms of bulk conditional means $\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\nu}(t) + \hat{\boldsymbol{\nu}}(t)$, we can rewrite it as in (5.5.15), i.e.

$$\begin{aligned}\mathcal{M}(t) =& \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{\nu}(t) + \boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{\nu}(t)) + \\ &\int_0^t dt'\,\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\left[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{\nu}(t')) + \boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t'))\right]\end{aligned} \tag{K.1.2}$$

with $\boldsymbol{\nu}(t') = \int_0^{t'} dt''[e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}]\delta\boldsymbol{x}^{\mathrm{s}}(t'')$, a vector of size N^{b} (number of bulk species). Some time integrals in the memory are taken into account by this substitution and an equivalent numerical implementation can be performed by solving (K.1.1) jointly with the N^{b} *linear* differential equations for $\boldsymbol{\nu}(t)$

$$\frac{d\boldsymbol{\nu}(t)}{dt} = \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{\nu}(t) + \boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t) \tag{K.1.3}$$

The time integral left in the last term of (K.1.2) appears as a result of the insertion of $\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)$ in the effective action. This term can be rearranged and evaluated by introducing *ad hoc* auxiliary variables. We need to decompose the exponential kernel $e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}$ into a superposition of pure

exponentials (by diagonalizing $\boldsymbol{K}^{\mathrm{bb}}$)

$$\begin{aligned}&\int_0^t dt' \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\left[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{\nu}(t'))+\boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t'))\right]=\\&=\int_0^t dt' \boldsymbol{K}^{\mathrm{sb}}\sum_{c=1}^{N^{\mathrm{b}}}\boldsymbol{r}^c\boldsymbol{l}^{c\,T}\,e^{\lambda_c(t-t')}\left[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{\nu}(t'))+\boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t'))\right]\end{aligned} \tag{K.1.4}$$

where λ_c are the N^{b} eigenvalues of $\boldsymbol{K}^{\mathrm{bb}}$ (with negative real part to ensure stability), $\boldsymbol{r}^c$ and $\boldsymbol{l}^c$ are respectively the right and left eigenvectors of $\boldsymbol{K}^{\mathrm{bb}}$, playing the role of the coefficients of this decomposition. We consider a vector $\boldsymbol{z}(t)=\{z^c(t)\}$, $c=1,...,N^{\mathrm{b}}$ whose components, one for each eigenvalue λ_c, satisfy

$$\frac{d}{dt}z^c(t)=\lambda_c\,z^c(t)+\boldsymbol{l}^{c\,T}\cdot\left[\boldsymbol{K}^{\mathrm{b,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{\nu}(t))+\boldsymbol{K}^{\mathrm{b,bb}}(\boldsymbol{\nu}(t)\circ\boldsymbol{\nu}(t))\right]\qquad z^c(0)=0 \tag{K.1.5}$$

then (K.1.1) can be translated into

$$\frac{d}{dt}\delta\boldsymbol{x}^{\mathrm{s}}(t)=\boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t)+\boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t))+\boldsymbol{K}^{\mathrm{sb}}\boldsymbol{\nu}(t)+\boldsymbol{K}^{\mathrm{s,sb}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{\nu}(t))+\boldsymbol{K}^{\mathrm{sb}}\sum_{c=1}^{N^{\mathrm{b}}}\boldsymbol{r}^c z^c(t) \tag{K.1.6}$$

As a conclusion, solving N^{s} subnetwork equations with integral memory terms is equivalent to solving a system with $2N^{\mathrm{b}}$ additional equations, the N^{b} ones describing $\boldsymbol{\nu}(t)$ (see (K.1.3)) and the N^{b} ones for $\boldsymbol{z}(t)$ (see (K.1.5)). For projection methods, the additional variables one needs to introduce for expressing memory integrals via differential equations is given by the dimension of the bulk variables space (see [77]) including all the possible products (thus scaling $\sim(N^{\mathrm{b}}\times N^{\mathrm{b}})+(N^{\mathrm{s}}\times N^{\mathrm{b}})$).

Chapter 6

Summary and Future directions

6.1 Summary

In this thesis, we have presented approaches inspired by theoretical physics to tackle model reduction and the inverse problem of state prediction in large networks of continuous degrees of freedom (d.o.f.), bearing in mind the necessity for these techniques in biochemical systems with noise. In particular, we applied path integrals to obtain effective Gaussian representations of Langevin-type dynamics and the construction of both the Extended Plefka Expansion (chapter 2) and the Gaussian Variational Approximation (chapter 5) can be rationalized from this point of view. This choice was motivated by the fact that the theory of Gaussian processes is well established [16] and Gaussian distributions (more generally exponential families) exhibit a "closure" property, i.e. marginals and conditionals are still Gaussian. In addition, the characterization of the dynamics consists of equations for just the first two moments as functions of time. Importantly, in deriving the Gaussian Variational Approximation, we keep the couplings among trajectories rather than decoupling them, in such a way as to retain the *microscopic* details of their joint temporal evolution; on the other hand, the Extended Plefka Expansion, where by construction we neglect the cross-correlations between species, actually provides a *mean field* type approximation. For the latter, the full dynamics of the network is assumed to contain arbitrary nonlinearities and interactions in a generic drift $\boldsymbol{\phi}(\boldsymbol{x}(t))$. The logic of the expansion can be summarized as follows. By appeal to the Martin–Siggia–Rose–Janssen–De Dominicis (MSRJD) path integral formalism [19–22] one obtains a representation of the dynamical trajectories $\boldsymbol{x}(t)$ comprising also auxiliary trajectories $\hat{\boldsymbol{x}}(t)$ which play the role of responses. In analogy with the Plefka expansion well known for statics [33], we introduce a parameter α to multiply $\boldsymbol{\phi}(\boldsymbol{x}(t))$, i.e. controlling the interaction strength, in such a way that $\alpha = 0$ is equivalent to the non-interacting dynamics. We define a Gibbs free energy by

introducing a set of fields acting as a "bias" of the path ensemble to constrain the first and second moments of the resulting probability distribution over the physical and auxiliary trajectories: the *extension* relies on the inclusion of these second moments when we deal with the dynamics of continuous d.o.f.. Next we expand the Gibbs energy for small α and up to the second order and setting the biases to zero allows us to recast the interactions in terms of fields that we call "effective" as they convey an effectively non-interacting version of the initial dynamics. As a result of writing the approximated dynamics in terms of these effective fields, the coupling between trajectories is replaced by a coupling of each trajectory to its own past via a memory and a coloured noise term. The non-linear self-consistent relations one can derive for means, correlations and responses, the order parameters of the theory, are the core of this new dynamical mean field description.

While the Extended Plefka Expansion is a novel contribution of this work, the Gaussian Variational Approximation, an already known approximation technique [110], is here revisited in the MSRJD path integral language. The starting point is a Langevin dynamics describing unary and binary reactions (i.e. biochemical complex formation and dissociation). By appeal to MSRJD path integrals, one can formally write the probability distribution of $\boldsymbol{x}(t)$ and $\hat{\boldsymbol{x}}(t)$ which would be nevertheless infeasible to evaluate exactly; thus one can approximate it with a Gaussian to be found variationally, i.e. by minimizing a pseudo-distance (the Kullback-Leibler divergence [114]) between the real and the approximating distribution. This approximate framework is chosen as convenient to work out the "reduced" dynamics by marginalization, i.e. by integrating over the degrees of freedom we are not interested in. Here "reduced" dynamics refers to the equations of motion for the subnetwork, the subset of elements of interest. One can imagine selecting these because better characterized either from the experimental or theoretical point of view. More precisely, we read the equations reduced to the subnetwork encoded by this Gaussian variational distribution and we find that they contain the reactions within the subnetwork in their original form, a memory term and an additional coloured noise. The memory accounts for the fact that the subnetwork interacts with the bulk, such effects propagate through the bulk and they feed back into the subnetwork dynamics with some delay. The coloured noise stems from the propagation of fluctuations and the uncertainty of the initial conditions in the bulk. We exploit a perturbative expansion about a linearized dynamics as an analytically controlled way for deriving the expressions for the memory function and the coloured noise covariance. We show explicitly that the resulting subnetwork description is equivalent to the one obtained by projection methods in [77] up to $O(\delta \boldsymbol{x}^2)$ and for negligible noise. Including the memory terms leads to higher quantitative accuracy of prediction compared to other standard approximation schemes (isolated subnetwork and bulk in steady state), as we illustrate for a toy model and for a subnetwork chosen from the protein-protein interaction

network around the Epidermal Growth Factor Receptor [116]. In addition, memory and coloured noise arising via the bulk, which can be seen as an embedding environment, provide a characterization of the extrinsic noise, whose importance in shaping biological dynamics is increasingly acknowledged [8, 150, 151].

We have furthermore argued the usefulness of path integral methods in the context of inference for dynamics. In chapter 3, we apply the Extended Plefka Expansion to study the Gaussian distribution over the bulk trajectories conditioned on some fixed observations, such as the data available experimentally: while the conditional mean provides the best estimate of the bulk dynamics, the equal time variance quantifies the uncertainty of this prediction, thus is a measure of the associated error. In particular we look at an analytically tractable scenario to study the average performance case: bulk and subnetwork evolving and interacting via a linear dynamics. We consider the stationary regime (where time translation invariance makes it convenient to work in the Fourier space), mean field couplings (all-to-all, weak and long-ranged) and the limit of an infinitely large bulk size. Under these conditions, the errors become site-independent self-consistently and equivalent to the average error, which assesses the quality of the prediction from the *macroscopic* point of view. We perform a scaling analysis of amplitude and time correlations of the average error in the vicinity of the critical points in the parameter space (i.e. where the error would diverge). The relevant parameters in this regard are $\alpha = N^{\mathrm{s}}/N^{\mathrm{b}}$, the ratio[1] between observed and hidden nodes, and other structural parameters such as γ (related to the bulk internal stability) and p (related to the relative weight of self-interactions and hidden-to-observed couplings). In chapter 4, we derive the *exact* average performance case, i.e. without approximation. We imagine rewriting the bulk posterior dynamics in terms of an "effective" drift (where by effective we mean conditioned on observations): its expression contains in fact the original bulk-bulk couplings and a Lagrange multiplier enforcing the constraints given by the observed values. By taking the stationary limit of the Kalman filter [63] (the exact algorithmic technique for posterior distributions of linear dynamics) we find that the posterior variance satisfies a Lyapunov-type equation containing the effective drift. The aim is then to evaluate this expression in the thermodynamic limit, given the assumption, as in chapter 3, that the matrix of couplings has Gaussian distributed elements scaling as $1/\sqrt{N^{\mathrm{b}}}$ and with arbitrary degree of symmetry. We average w.r.t. to spectral densities known by Random Matrix Theory in the case of an equilibrium dynamics and by appeal to functional methods for the non-equilibrium case: we obtain the same closed form algebraic equation for the posterior

[1]Note that, in the inference chapters 3 and 4, α is used mainly to designate this ratio, not as in the previous chapters the small parameter for the Plefka expansion.

variance as from the extended Plefka expansion, which can therefore be said to become exact in the thermodynamic limit for a linear dynamics. This equivalence, an essential and non-trivial consistency check for this novel framework, can be fully verified also for the case without observations, as developed in chapter 2, and its validity ultimately relies on the fact that interaction terms become Gaussian for large networks with mean field couplings (from the central limit theorem).

As is clear from this work, a path integral representation of dynamics has a structure which is amenable to the application of variational principles and convenient for formal manipulations (e.g. mean field approximations, marginalization over trajectories); furthermore it offers a flexible framework to develop a perturbative theory for nonlinear problems. More generally, mathematical tools borrowed from field theories in physics, such as path integrals (see also [136, 166]) and Feynman diagrams [160], yield more accurate prediction methods and an understanding of dynamical behaviours grounded on detailed descriptions of fundamental components. These results are promising and could motivate a systematic transfer of approaches from theoretical physics to systems biology, still to be fully explored.

6.2 Future directions

Several aspects of the path integral approaches we have presented in this thesis deserve further investigation, including the other applications and comparisons that we have already sketched in the discussion at the end of each chapter. Let us here summarize and elaborate further how our results could be expanded and bring future contributions to the fields of statistical physics, systems biology and inference methods.

6.2.1 Implementation and performance

It would be interesting to assess more systematically the performance of the extended Plefka expansion in the presence of nonlinearities in the dynamics and on different types of graphs. For example here (chapters 2 and 3) we considered fully connected networks described by Gaussian, long-ranged interactions, whose behaviour is properly mean field; the performance of the extended Plefka expansion should be maximal on these networks but an open question is to what extent it would be still be reliable on non mean field networks, e.g. power law networks. An idea for future work is then to implement this approximation for particular models and to compare the outcome with the baseline of direct simulation of the system, across a range of parameter values. It will be interesting to explore especially the dependence of its performance on the system size, the

amount of dynamical noise, the strength and the symmetry of interactions, and the choice of initial conditions.

6.2.2 Coarse-graining fluctuations in gene expression

A crucial objective, on which a systematic validation of our methods relies, is the application to biological complex systems. We have developed the analysis of chapter 5 for unary and binary reactions and provided examples of implementation in biochemical networks, but we aim to explore biological applications in more detail. Indeed one still needs to thoroughly investigate the potential of our theoretical frameworks, e.g. by designing extensions or variants tailored to different settings. One could hope that they can be successfully utilized not only to reduce network complexity but also to shed light on mechanisms not fully understood, such as the source and the propagation of fluctuations in gene regulation.

Signal-transduction networks (built up of interacting proteins) operate at typical rates faster than gene expression, which can be assumed to be approximately at steady state given the slow timescale. This observation justifies our initial approach, focussing for simplicity on proteome dynamics (see chapter 5). As a future goal, one could start to integrate the gene expression level into the protein interaction network and use our setup to analyze regulation mechanisms. Gene expression consists of the transcription of DNA to messenger RNA (mRNA) followed by translation to proteins in the cytoplasm. It is generally acknowledged as a stochastic process, due to the low copy number of DNA and mRNA molecules involved, and all the participating processes along the regulatory chain propagate rather than average out fluctuations. A potential way to keep track of how the randomness at gene expression level affects fluctuations in protein abundances is a coarse-graining procedure, whose typical domain of application are systems with different temporal scales. In this spirit, descriptions of events at a particular scale are improved by accounting for "effective" terms stemming from an integration over processes at either finer or coarser scales. An interesting recent observation is that proteins are produced by releasing large bursts during short time intervals and each burst originates from stochastic events during transcription and translation [167]. In light of this, Pedraza and Paulsson [167] estimate the stationary variance in protein abundance by adding to the copy number noise a coarse-grained term that describes the noise inherited from mRNA and depends on bursts and waiting-time statistics. The basic assumption is that protein levels do not immediately adjust to changes in the protein synthesis rate, thus the measured values effectively are determined by an average over mRNA fluctuations in a defined temporal interval according to an appropriately chosen statistical distribution. The underlying molecular mechanisms are not

sufficiently understood to predict specifically the statistics of burst size and frequency, which crucially reflect the dynamics of how fluctuations are generated. Different hypotheses can be nevertheless put forward and validated through a coarse-grained approach, which could then provide more conclusive evidence by direct comparison of its predictions with experimental findings on protein abundances. Estimating such coarse-grained contributions is therefore an essential step to integrate processes at the gene expression and protein level: the model reduction strategy we have presented in chapter 5 could be applied to achieve this.

6.2.3 Learning rules and experiment design

The results of this work can help deal with partially observed networks, in particular to draw some inference on the uncharacterized part (the bulk) starting from just subnetwork equations. This is an essential task in systems biology, since noise and missing variables typically prevent one from a complete knowledge of the system. So far we have focussed on state inference, which covers only one aspect of what in general are referred to as *inverse* problems. A future direction would then be to address, by using the same approach, other complementary questions, such as learning of unknown interactions and parameter fitting as well as identifiability[2]. Optimizing the data likelihood in chapter 3, calculated in terms of path integrals, with respect to the couplings would by definition give a maximum-likelihood estimate for these quantities. The relevant learning rules could be developed starting from the means/correlations equations for fixed couplings. One could investigate aspects such as the accuracy of the inferred couplings and the computational efficiency of the iterative algorithms for implementing the learning rules. Importantly, as discussed in chapter 3, by accounting for temporal series (which might be available from experiments) as paths, we are actually incorporating additional information stemming from the *temporal structure* of data: it would be interesting to assess more systematically the advantages of modelling data explicitly as a trajectory over time. It could be relevant for parameter fitting, thus *model selection*, as it can add constraints to the estimation. Many models are in fact compatible with a set of parameters and it is then important to assess to what extent this range of different values is due to experimental uncertainty or reveals something more fundamental such as *robustness* of biological functions.
Let us stress that our results could be already relevant in this regard, as tackling inverse problems relies on an interplay between state inference and parameter estimation. In this thesis, we have analyzed the inference problem for the time courses of hidden nodes, assuming that the model

[2]Parameters identifiability is the problem of what parameters can be inferred: Fisher Information Matrix (FIM) is a classical criterion used in this regard, see [168].

parameters (couplings) are randomly distributed with known average properties (degree of symmetry, strength). As a next step, one could think of considering the parameters to be determined by Expectation-Maximization algorithm [54, 169], where the Expectation part relies precisely on computing the posterior statistics. Our simple expressions for the average posterior variance in terms of the average coupling strength and degree of symmetry would then simplify this procedure in some regimes and help investigate it in an analytically controlled, thus more insightful, way.

Moreover, our results could be of interest in *experiment design* when the spatio-temporal evolution of just a few species is actually under control. Which and how many variables one should measure, which species should be perturbed to get information on other ones, how reliable is the prediction depending on the amount of data and known properties of the system: these are all questions of theoretical and practical significance towards the choice of experimental settings as well as the interpretation of results. In chapters 3 and 4, we analyzed how the inference error depends on structural parameters of the system such as the ratio between the numbers of observed and unknown nodes of the network, the degree of symmetry of the interactions and their amplitudes relative to the decay constant of the internal unknown dynamics: all of them can be considered as measurable at least indirectly. One could think of settings in which such a study might actually guide the experiment preparation in such a way to maximize the inference accuracy. For example the parameter α, measuring the relative number of nodes to monitor over time, could enter the specification of a hypothetic experimental protocol. If we suppose that an estimate of other parameters (γ, p, η) is available either from previous measurements or some a priori knowledge, explicit expressions for $C(0)$, the average inference error when many hidden trajectories are reconstructed, in some region of the phase space might serve to fix a minimal α for $C(0)$ to be below certain set precision thresholds. We could also think of extending the framework presented in chapter 5 and incorporating a systematic analysis of the response to perturbations, possibly with the addition of gene expression. A similar analysis has been developed with projection methods to predict the change in steady states and memory functions expected in the subnetwork from perturbations in the bulk [134]. In this way one obtains results directly comparable with experiments of gene knockdowns and targeted perturbations of individual components of signalling pathways [111, 170]. Importantly, they could help choose which perturbations should be applied to best deduce unknown values of kinetic parameters and/or concentrations.

6.2.4 Connection to information-theoretic tools

There are precedents for the use of path integral approaches to develop sensitivity estimators [171, 172] and tools of fidelity control in coarse-grained descriptions [173], thus tackling questions related to uncertainty quantification, model reduction and model selection. The main idea is that path distributions allow one to define the dynamical analogues of the Relative Entropy Rate (RER) - whose definition corresponds to the KL divergence - or Fisher Information Matrix (FIM) for systems not satisfying equilibrium assumptions. In these existing approaches based on path integrals and concepts from information theory one sees at work the following core ideas: measuring the loss of information by defining pseudo-distances (RER) in the path space and finding the most faithful coarse-grained models or the most robust set of parameters by minimizing such a distance. Similarly, in chapter 5, the KL minimization allows us to retrieve the optimally parametrized description in terms of Gaussian moments, i.e. the means and correlations defining the approximating distribution for the whole temporal trajectory.

In the case of sensitivity analysis, the RER between path distributions quantifies the information loss/change in changing from set of parameters θ to a perturbed one $\theta + \epsilon$. Consistently with this, the FIM is used as a methodology for analyzing the sensitivity of path distributions with respect to the perturbation ϵ in the parameters (due to uncertainty or experimental errors). In the path space, information-theoretic tools, such as RER and the KL divergence, and FIM can be treated as observables of the stochastic dynamics, thus estimated by Monte Carlo methods. Sampling them is necessary when simpler approximations are not applicable, e.g. when one expects large deviations or non-Gaussianity in the long time behaviour such as for sustained random oscillations [174]. In some regimes, such as well mixing species which reach a single steady state, the RER and FIM can be calculated analytically: e.g. in [168] the authors apply the Linear Noise Approximation (LNA) [64] for this purpose in such a way as to reduce the FIM calculation to a set of ODEs. (The LNA has a Gaussian solution, thus once equations for means, variances and correlations are given, deriving them with respect to parameters leads one to the equations for the FIM). One could follow a similar baseline and estimate the FIM in our Gaussian reduced framework (chapter 5). In this way we could study the sensitivity w.r.t. variation in parameters in the full network starting from just subnetwork equations; this would contribute to model selection strategies when only a few species can be controlled externally. More generally, as potential research directions, one could think of defining information-theoretic quantities within our frameworks to see whether parameter estimation and sensitivity analysis could be addressed in an efficient way.

Malheureusement, les choses ne se passent pas avec cette simplicité. Il se peut bien - et encore ce serait à examiner - qu'une science naisse d'une autre mais jamais une science ne peut naître de l'absence d'une autre, ni de l'échec, ni même de l'obstacle rencontré par une autre. [3]

Michel Foucault, "Les mots et les choses"

Multa novis verbis praesertim cum sit agendum
propter egestatem linguae et rerum novitatem [4]

Lucretius, "De rerum natura", I, 138-139

[3]Unfortunately, things do not happen as simply as that. It is quite possible - though it would be a matter requiring careful scrutiny - that one science can arise out of another; but no science can be generated by the absence of another, or from another's failure, or even from some obstacle another has encountered.

[4]Above all as many things must be treated in new words, because of the poverty of our tongue and the newness of the themes.

References

[1] J. Ackermann, J. Einloft, J. Nothen, and I. Koch. Reduction techniques for network validation in systems biology. *J. Theor. Biol.*, 315:71–80, 2012.

[2] O. Radulescu, A.N. Gorban, A. Zinovyev, and V. Noel. Reduction of dynamical biochemical reactions networks in computational biology. *Front. Genet.*, 3(131), 2012.

[3] C. W. Gardiner. Adiabatic elimination in stochastic systems: formulation of methods and applications to few-variable systems. *Phys. Rev A*, 29(5):2814–2822, 1984.

[4] P. Thomas, A.V. Straube, and R. Grima. The Slow-Scale Linear Noise Approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Syst. Biol.*, 6(39), 2012.

[5] M. Apri, M. De Gee, and J. Molenaar. Complexity reduction preserving dynamical behavior of biochemical networks. *J. Theor. Biol.*, 304:16–26, 2012.

[6] M.S. Okino and M.L. Mavrovouniotis. Simplification of mathematical models of chemical reaction systems. *Chem. Rev.*, 98(2):391–408, 1998.

[7] M. Sunnaker, G. Cedersund, and M. Jirstrand. A method for zooming of nonlinear models of biochemical systems. *BMC Systems Biology*, 5(140), 2011.

[8] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[9] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135, 2008.

[10] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 3rd Edition, 2007.

[11] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, (22):403–434, 1976.

[12] D. F. Anderson and T. G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In H. Koeppl et al., editor, *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*, chapter 1, pages 1–44. Springer, 2011.

[13] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(044104), 2006.

[14] S. Peles, B. Munsky, and M. Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *J. Chem. Phys.*, 125 (204104), 2006.

[15] S. Smith, C. Cianci, and R. Grima. Model reduction for stochastic chemical systems with abundant species. *J. Chem. Phys.*, 143(214105), 2015.

[16] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[17] N. Wiener. *Collected works: with commentaries*. MIT Press, 1976.

[18] L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Phys. Rev.*, 21(6): 1505–1521, 1953.

[19] P.C. Martin, E.D. Siggia, and H.A. Rose. Statistical dynamics of classical systems. *Phys. Rev. A*, 8:423–436, 1973.

[20] C. De Dominicis. Dynamics as a substitute for replicas in systens with quenched random impurities. *Phys. Rev. B*, 18(9):4913–4919, 1978.

[21] H. K. Janssen. Lagrangian for classical field dynamics and renormalization group calculations of dynamical critical properties. *Z. Phys. B: Cond. Mat.*, 23:377–380, 1976.

[22] C. De Dominicis and L. Peliti. Field-theory renormalization and critical dynamics above T_c: Helium, antiferromagnets, and liquid-gas systems. *Phys. Rev. B*, 18(1):353–375, 1978.

[23] J. A. Hertz, Y. Roudi, and P. Sollich. Path integral methods for the dynamics of stochastic and disordered systems. *Arxiv preprint 1604.05775*, 2016.

[24] C. C. Chow and M. A. Buice. Path integral methods for stochastic differential equations. *JMN*, 5(8), 2015.

[25] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. World scientific publishing, 4th edition, 2006.

[26] B. Bravi, P. Sollich, and M. Opper. Extended Plefka expansion for stochastic dynamics. *J. Phys. A: Math. Theor.*, 49(194003), 2016.

[27] C. W. Gardiner. *Handbook for Stochastic Methods*. Springer, 2nd edition, 1985.

[28] Z. Szallasi, J. Stelling, and V. Periwal, editors. *System Modeling in Cellular Biology*. MIT press, 2006.

[29] F. Achcar, E. J. Kerkhoven, The SilicoTryp Consortium, B. M. Bakker, M. P. Barrett, and R. Breitling. Dynamic modelling under uncertainty: the case of *Trypanosoma brucei* energy metabolism. *PLoS Comput. Biol.*, 8(I):1–11, 2012.

[30] M. Mezard, G. Parisi, and M.A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, 1987.

[31] M. Opper and D. Saad, editors. *Advanced Mean Field Methods*. MIT press, 2001.

[32] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35(46): 1792–1796, 1975.

[33] T. Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A: Math. Gen.*, 15:1971–1978, 1982.

[34] D.J. Thouless, P.W. Anderson, and R.G. Palmer. Solution of "Solvable model of a spin glass". *Phil. Mag.*, 35:593–601, 1977.

[35] Y. Roudi and J. Hertz. Dynamical TAP equations for non-equilibrium Ising spin glasses. *J. Stat. Mech.*, P03031, 2011.

[36] M.L. Mehta. *Random Matrices*. Elsevier-Academic Press, Amsterdam, 3rd edition, 2004.

[37] L. Bachschmid-Romano, C. Battistin, M. Opper, and Y. Roudi. Variational perturbation and extended Plefka approaches to dynamics on random networks: the case of the kinetic Ising model. *Arxiv preprint 1607.08379*, 2016.

[38] A. C. C. Coolen. *Handbook of Biological Physics*, volume 4, chapter Statistical Mechanics of Recurrent Neural Networks II Dynamics. Elsevier-Academic Press, Amsterdam, 2001.

[39] H. Eissfeller and M. Opper. Mean-field Monte Carlo approach to the Sherrington-Kirkpatrick model with asymmetric couplings. *Phys. Rev. E*, 50:709–720, 1994.

[40] L. F. Cugliandolo and D. S. Dean. Full dynamical solution for a spherical spin-glass model. *J. Phys. A: Math. Gen.*, 28:4213–4234, 1995.

[41] V. L. Girko. Circular law. *Theory Probab. Appl.*, 29:694–706, 1984.

[42] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.*, 6:440–449, 1965.

[43] V. L. Girko. Elliptic law. *Theory Probab. Appl.*, 30:677–690, 1986.

[44] B. Mehlig and J.T. Chalker. Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles. *J. Math. Phys.*, 41:3233, 2000.

[45] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.*, 60:1895–1899, 1988.

[46] H. B. Callen and T. A. Welton. Irreversibility and generalized noise. *Phys. Rev.*, 83:34–40, 1951.

[47] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Washington, 10th printing edition, 1972.

[48] R. Chetrite, G. Falkovich, and K. Gawedzki. Fluctuation relations in simple examples of non-equilibrium steady states. *J. Stat. Mech.*, P08005, 2008.

[49] H. Sompolinsky and A. Zippelius. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B*, 25(11):6860–6875, 1982.

[50] A. Crisanti and H. Sompolinsky. Dynamics of spin systems with random asymmetric bonds: Langevin dynamics and a spherical model. *Phys. Rev. A*, 36(10):4922–4939, 1987.

[51] H. Sompolinsky, A. Crisanti, and H.J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61(3):259–262, 1988.

[52] G. Biroli. Dynamical TAP approach to mean field glassy systems. *J. Phys. A: Math. Gen.*, 32(48):8365–8388, 1999.

[53] T. R. Kirkpatrick and D. Thirumalai. p-spin-interaction spin glass models: Connections with the structural glass problem. *Phys. Rev. B*, 36(10):5388–5397, 1987.

[54] T. P. Minka. Expectation Propagation for approximate Bayesian inference. *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI)*, pages 362–369, 2001.

[55] M. Opper and O. Winther. Expectation consistent approximate inference. *JMLR*, 6:2177–2204, 2005.

[56] B. Bravi and P. Sollich. Inference for dynamics of continuous variables: the Extended Plefka Expansion with hidden nodes. *JSTAT (under review), Arxiv preprint 1603.05538*, 2016.

[57] L. Bachschmid-Romano and M. Opper. Inferring hidden states in a random kinetic Ising model: replica analysis. *J. Stat. Mech.*, (P06013), 2014.

[58] C. Battistin, J. Hertz, J. Tyrcha, and Y. Roudi. Belief-propagation and replicas for inference and learning in a kinetic Ising model with hidden spins. *J. Stat. Mech.*, (P05021), 2015.

[59] B. Dunn and Y. Roudi. Learning and inference in a nonequilibrium Ising model with hidden spins. *Phys. Rev. E*, 87(022127), 2013.

[60] D. Trejo-Banos, A. J. Millar, and G. Sanguinetti. Experimental design for inference over *A. Thaliana* circadian clock network. *Computational Methods in Systems Biology (CMSB)*, 2015.

[61] P. J. M. Jones, A. Sim, H. B. Taylor, L. Bugeon, M. J. Dallman, B. Pereira, M. P. H. Stumpf, and J. Liepe. Inference of random walk models to describe leukocyte migration. *Phys. Biol.*, (12):1–12, 2015.

[62] F. Buettner, K. N. Natarajan, F. P. Casale, V. Proserpio, A. Scialdone, F. J. Theis, S. A. Teichmann, J. C. Marioni, and O. Stegle. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnol.*, 33(2):155–160, 2015.

[63] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82(1):35–45, 1960.

[64] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the Linear Noise Approximation. *Genome Res.*, 13(11):2475–2484, 2003.

[65] J. A. Hertz, A. Krogh, and G. I. Thorbergsson. Phase transitions in simple learning. *J. Phys. A: Math. Gen.*, (22):2133–2150, 1989.

[66] M. Opper. Learning in neural networks: Solvable dynamics. *Europhys. Lett.*, 8(4):389–392, 1989.

[67] D. De Martino, F. Capuani, M. Mori, A. De Martino, and E. Marinari. Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks. *Metabolites*, (3), 2013.

[68] P. Holme. Detecting degree symmetries in networks. *Phys. Rev. E*, (74):036107, 2006.

[69] J. Tyrcha and J. Hertz. Network inference with hidden units. *MBE*, 11(1):149–165, 2014.

[70] G. O. Fruhwirth, L. P. Fernandes, G. Weitsman, G. Patel, M. Kelleher, K. Lawler, A. Brock, S. P. Poland, D. R. Matthews, G. Keri, P. R. Barber, B. J. Vojnovic, S. M. Ameer-Beg, A. C. C. Coolen, F. Fraternali, and T. Ng. How Förster Resonance Energy Transfer imaging improves the understanding of protein interaction networks in cancer biology. *Chem. Phys.*, 12:442–461, 2011.

[71] N. G. Van Kampen. Itô versus Stratonovich. *J. Stat. Phys.*, 24(1):175–187, 1982.

[72] H. J. Kappen and J. J. Spanjers. Mean field theory for asymmetric neural networks. *Phys. Rev. E*, 61(5), 2000.

[73] M. Mezard and J. Sakellariou. Exact mean field inference in asymmetric kinetic Ising systems. *J. Stat. Mech.*, (L07001), 2011.

[74] R. L. Stratonovich. Conditional Markov processes. *Theor. Probab. Appl.*, 5(2):156–178, 1960.

[75] A. Kamenev. *Field theory of non-equilibrium systems*. Cambridge University Press, 2011.

[76] J. Cardy. *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, 1996.

[77] K. J. Rubin, K. Lawler, P. Sollich, and T. Ng. Memory effects in biochemical networks as the natural counterpart of extrinsic noise. *J. Theor. Biol.*, 357:245–267, 2014.

[78] L. Bachschmid-Romano and M. Opper. Learning of couplings for random asymmetric kinetic Ising models revisited: random correlation matrices and learning curves. *J. Stat. Mech.*, (P09016), 2015.

[79] A. Krogh. Learning with noise in a linear perceptron. *J. Phys. A. Math. Gen.*, 25:1119–1133, 1992.

[80] M. Opper and G. Sanguinetti. Variational inference for Markov jump processes. *Adv. Neural Inf. Process. Syst.*, 2007.

[81] M. Opper, A. Ruttor, and G. Sanguinetti. Approximate inference in continuous time Gaussian-jump processes. *Adv. Neural Inf. Process. Syst.*, 23:1831–1839, 2010.

[82] B. Cseke, M. Opper, and G. Sanguinetti. Approximate inference in latent Gaussian-Markov models from continuous time observations. *Adv. Neural Inf. Process. Syst.*, 26:971–979, 2013.

[83] P. Sollich. Finite-size effects in learning and generalization in linear perceptrons. *J. Phys. A. Math. Gen.*, 27:7771–7784, 1994.

[84] B. Bravi, M. Opper, and P. Sollich. Inferring the hidden states of a Langevin dynamics on large networks: the average case performance. *Phys. Rev. E (accepted), Arxiv preprint 1607.01622*, 2016.

[85] A. Braunstein, A. Pagnani, M. Weigt, and R. Zecchina. Inference algorithms for gene networks: A statistical-mechanics analysis. *JSTAT*, page P12001, 2008.

[86] R. S. Tsay. *Analysis of Financial Time Series, 3rd edition*. Wiley, 2010.

[87] A. S. Cofiño, J. M. Gutiérrez, B. Jakubiak, and M. Melonek. Implementation of data mining techniques for meteorological applications. *Realizing Teracomputing*, W. Zwieflhofer and N. Kreitz (Eds.), World Scientific:215–140, 2013.

[88] A. Engel and C. Van den Broeck. *Statistical Mechanics of Learning*. Cambridge University Press, 2004.

[89] W. Kinzel and M. Opper. *Models of Neural Networks III*, chapter Statistical Mechanics of Generalization. Springer, 1996.

[90] H. Sompolinsky, N. Tishby, and H. S. Seung. Learning from examples in large neural networks. *Phys. Rev. Lett.*, 65(13):1683–1687, 1990.

[91] Y. Le Cun, I. Kanter, and S. A. Solla. Eigenvalues of covariance matrices: Application to neural-network learning. *Phys. Rev. Lett.*, 66(18):2396–2399, 1991.

[92] S. F. Edwards and R. C. Jones. The eigenvalue spectrum of a large symmetric random matrix. *J. Phys. A: Math. Gen.*, 9(10):1595–1603, 1976.

[93] Y. Roudi and J. Hertz. Mean field theory for nonequilibrium network reconstruction. *Phys. Rev. Lett.*, 106(048702), 2011.

[94] J. Berg. Out-of-equilibrium dynamics of gene expression and the Jarzynski equality. *Phys. Rev. Lett.*, 18(100):188101–188105, 2008.

[95] M. Opper and G. Sanguinetti. Learning combinatorial transcriptional dynamics from gene expression data. *Bioinformatics*, 26(13):1623–1629, 2010.

[96] M. H. A. Davis and R. B. Vinter. *Stochastic Modelling and Control*. Chapman and Hall, 1985.

[97] A. Vakili and B. Hassibi. On the asymptotic eigenvalue distribution of certain random Lyapunov and Riccati recursions. *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 453–458, 2008.

[98] J. Bun, R. Allez, J. P. Bouchaud, and M. Potters. Rotational invariant estimator for general noisy matrices. *Arxiv preprint 1502:06736*, 2015.

[99] A. Ruttor and M. Opper. Efficient statistical inference for stochastic reaction processes. *Phys. Rev. Lett.*, 103(23):230601, 2009.

[100] P. Erdős and A. Rényi. On random graphs I. *Publicationes Mathematicae*, 6(290-297), 1959.

[101] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina. Bayesian inference of epidemics on networks via Belief Propagation. *Phys. Rev. Lett.*, 112(11): 118701, 2014.

[102] J. Bindi, A. Braunstein, and L. Dall'Asta. Predicting epidemic evolution on contact networks from partial observations. *Arxiv pre-print 1608.06516*, 2016.

[103] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-sb*, 1(457), 1967.

[104] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume I of *CRM Monograph Series*. AMS, 1996.

[105] R. Speicher. Free probability and random matrices. *Proceedings of the ICM*, III:477–501, 2014.

[106] Z. Burda. Free products of large random matrices- a short review of recent developments. *J. Phys.: Conf. Ser.*, 473(012002), 2013.

[107] E. Domany, J. L. van Hemmen, and K. Schulten, editors. *Models of Neural Networks III*. Springer, 1996.

[108] R. Kühn. Spectra of sparse random matrices. *J. Phys. A: Math. Theor.*, 41(295002), 2008.

[109] T. Rogers, I. Pérez Castillo, R. Kühn, and K. Takeda. Cavity approach to the spectral density of sparse symmetric random matrices. *Phys. Rev. E*, 78(031116), 2008.

[110] C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor. Gaussian process approximations of stochastic differential equations. *JMLR: Workshop and Conference Proceedings*, 1:1–16, 2007.

[111] Molinelli E. J. et al. Perturbation biology: Inferring signaling networks in cellular systems. *PLOS Computational Biology*, 2013.

[112] L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, 1989.

[113] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Oxford University Press, 1995.

[114] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22(1): 79–86, 1951.

[115] B. Bravi and P. Sollich. Statistical physics approaches to subnetwork dynamics in biochemical systems. *Phys. Biol. (to be submitted)*, 2016.

[116] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.*, 274(42):30169–30181, 1999.

[117] A. Plotnikov, E. Zehorai, S. Procaccia, and R. Seger. The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et Biophysica Acta*, 1813:1619–1633, 2011.

[118] G.L. Eyink. Action principle in nonequilibrium statistical dynamics. *Phys. Rev. E*, 54(4), 1996.

[119] M. Sasai and P.G. Wolynes. Stochastic gene expression as many body problem. *PNAS*, 100 (5):2374–2379, 2003.

[120] J. Ohkubo. A field theoretic approach to master equations and a variational method beyond the Poisson ansatz. *JSTAT*, P09017, 2007.

[121] I. Cohn, T. El-Hay, N. Friedman, and R. Kupferman. Mean field variational approximation for continuous-time Bayesian networks. *Proceedings of NIPS*, 2009.

[122] P. Whittle. On the use of the normal approximation in the treatment of stochastic processes. *J.R. Stat. Soc.*, 19(2):268–281, 1957.

[123] C. Bosia, A. Pagnani, and R. Zecchina. Modelling competing endogenous RNA networks. *Plos One*, 8(6), 2013.

[124] E. Lakatos, A. Ale, P. D. W. Kirk, and M. P. H. Stumpf. Multivariate moment closure techniques for stochastic kinetic models. *J. Chem. Phys.*, 143(094107), 2015.

[125] D. Schnoerr, G. Sanguinetti, and R. Grima. Validity conditions for moment closure approximations in stochastic chemical kinetics. *J. Chem. Phys.*, 141(084103), 2014.

[126] D. Schnoerr, G. Sanguinetti, and R. Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.*, 143(185101), 2015.

[127] R. Grima. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.*, 136(154105), 2012.

[128] R. Zwanzig. Memory effects in irreversible thermodynamics. *Phys. Rev.*, 124(4), 1961.

[129] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *PNAS*, 97(7):2968–2973, 2000.

[130] A. J. Chorin, O. H. Hald, and R. Kupferman. Prediction from partial data, renormalization, and averaging. *J. Scient. Comput.*, 28(2/3), 2006.

[131] P. Thomas, R. Grima, and A.V. Straube. Rigorous elimination of fast stochastic variables from the Linear Noise Approximation using projection operators. *Phys. Rev. E*, 86:041110, 2012.

[132] D. Cubero and S.N. Yaliraki. Inhomogeneous multiscale dynamics in harmonic lattices. *J. Chem. Phys.*, 122(034108), 2005.

[133] P. Espanol. Coarse graining from coarse-grained descriptions. *Phil. Trans. R. Soc. Lond.*, 360:383–394, 2002.

[134] K. Rubin. *Dynamics of protein interaction subnetworks*. PhD thesis, King's College London, 2014.

[135] K. J. Rubin and P. Sollich. Michaelis-Menten dynamics in protein subnetworks. *J. Chem. Phys.*, 144(174114), 2016.

[136] N.A. Sinitsyn, N. Hengartner, I. Nemenman, and W.H. Press. Adiabatic coarse-graining and simulations of stochastic biochemical networks. *PNAS*, 106(26):10546–10551, 2009.

[137] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11(2):197–202, 2014.

[138] C. Zechner and H. Koeppl. Uncoupled analysis of stochastic reaction networks in fluctuating environments. *PLoS Comput. Biol.*, 10(12):e1003942, 2014.

[139] A.C.C. Coolen and S. Rabello. Generating functional analysis of complex formation and dissociation in large protein interaction networks. *JPCS*, 197, 2009.

[140] F. Ritort and P. Sollich. Glassy dynamics of kinetically constrained models. *Advances in Physics*, 52(4):219 — 342, 2003.

[141] R. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys.*, 9(3), 1973.

[142] H. Mori. Transport, collective motion, and brownian motion. *Prog. Th. Phys.*, 33(3), 1965.

[143] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 166:239–257, 2002.

[144] K. H. Fischer and J. A. Hertz. *Spin Glasses*. Cambridge Studies in Magnetism, 1991.

[145] R. Grima, P. Thomas, and A.V. Straube. How accurate are the nonlinear Chemical Fokker-Planck and Chemical Langevin equations? *J. Chem. Phys.*, 135(084103), 2011.

[146] M.B. Diaz, R. Desikan, and M. Barahona. Linear models of activation cascades: analytical solutions and coarse-graining of delayed signal transduction. *Arxiv preprint 1112.0270v2*, 2015.

[147] N. Herath, A. Hamadeh, and D. Del Vecchio. Model reduction for a class of singularly perturbed stochastic differential equations. *Proceedings of ACC*, 2015.

[148] D. Saad and A. Mozeika. Emergence of equilibrium-like domains within non-equilibrium Ising spin dynamics. *Phys. Rev. E*, 87(3):032131, 2013.

[149] M. Barahona, A.C. Doherty, M. Sznaier, H. Mabuchi, and J.C. Doyle. Finite horizon model reduction and the appearance of dissipation on Hamiltonian systems. *Proceedings of the 41st IEEE Conference on Decision and Control*, 2002.

[150] P. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*, 99(20):12795–12800, 2002.

[151] V. Shahrezaei, J. F. Olivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4(196), 2008.

[152] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *PNAS*, 108(29):12167–12172, 2011.

[153] C.G. Bowsher and P.S. Swain. Identifying sources of variation and the flow of information in biochemical networks. *PNAS*, 109(20):E1320–E1328, 2012.

[154] B. Bravi and G. Longo. The unconventionality of nature: Biology, from noise to functional randomness. *Unconventional Computation & Natural Computation Conference (UCNC) Springer proceedings*, C. Calude and M.J. Dinneen, eds.:3–34, 2015.

[155] D. Holcman and Z. Schuss. Stochastic chemical reactions in microdomains. *J. Chem. Phys.*, 122(11470), 2005.

[156] T. Martin, X. Zhang, and M.E.J. Newman. Localization and centrality in networks. *Phys. Rev. E*, 90(052808):e1003942, 2014.

[157] R.R. Nadakuditi and M.E.J. Newman. Spectra of random graphs wih arbitrary expected degrees. *Phys. Rev. E*, 87(012803), 2013.

[158] M.T. Schaub, Y.N. Billeh, C.A. Anastassiou, C. Koch, and M. Barahona. Emergence of slow-switching assemblies in structured neuronal networks. *PLoS Comput. Biol.*, 11(7): e1004196, 2015.

[159] H.H. Chang, M. Hemberg, M. Barahona, D.E. Ingber, and S. Huang. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(06965):544–547, 2008.

[160] P. Thomas, C. Fleck, R. Grima, and N. Popovic. System size expansion using Feynman rules and diagrams. *J. Phys. A: Math. Theor.*, 47(455007), 2014.

[161] M. Doi. Stochastic theory of diffusion-controlled reaction. *J . Phys. A: Math. Gen.*, 9(9), 1976.

[162] M. Doi. Second quantization representation for classical many-particle system. *J . Phys. A: Math. Gen.*, 9(9), 1976.

[163] L. Peliti. Path integral approach to birth-death processes on a lattice. *Journal de Physique*, 46(9):1469–1483, 1985.

[164] N. Popovic, C. Marr, and P.S. Swain. A geometric analysis of fast-slow models for stochastic gene expression. *J. Math. Biol.*, 72:87–122, 2016.

[165] M. Opper and C. Archambeau. The variational gaussian approximation revisited. *Neural Comput.*, 21(3):786–92, 2009.

[166] F. A. N. Santos, H. Gadelha, and E. A. Gaffney. Fock space, symbolic algebra, and analytical solutions for small stochastic systems. *Phys. Rev. E*, 92:062714, 2015.

[167] J. M. Pedraza and J. Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, (319):339–343, 2008.

[168] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf. Sensitivity, robustness and identifiability in stochastic chemical kinetics models. *PNAS*, 108(21):8645–8650, 2011.

[169] G. J. McLachlan and T. Krishnan. *The EM Algorithm and its Extensions*. Wiley, 1997.

[170] N. Shayeghi, T. Ng, and A. C. C. Coolen. Direct Response Analysis in cellular signalling networks. *J. Theor. Biol.*, 304:219–225, 2012.

[171] Y. Pantazis and M. A. Katsoulakis. A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics. *J. Chem. Phys.*, 138:054115, 2013.

[172] A. Tsourtis, Y. Pantazis, M. A. Katsoulakis, and V. Harmandaris. Parametric sensitivity analysis for stochastic molecular systems using information theoretic metrics. *J. Chem. Phys.*, 143:014116, 2015.

[173] M. A. Katsoulakis and P. Plecháč. Information-theoretic tools for parametrized coarse-graining of non-equilibrium extended systems. *J. Chem. Phys.*, 139:074115, 2013.

[174] Y. Pantazis, M. A. Katsoulakis, and D. G. Vlachos. Parametric sensitivity analysis for biochemical reaction networks based on pathwise information theory. *BMC Bioinformatics*, 14(311), 2013.